



PD 30-OCT-1997.  
 XX  
 PF 17-APR-1997; 97WO-US06345.  
 XX  
 PR 19-APR-1996; 96US-0635130.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Caras IW;  
 XX  
 DR WPI, 1997-535837/49.  
 XX  
 DR N-PSDB; AAV06354.  
 XX  
 PT Human AL-2 neurotrophic factor and related DNA - used to develop  
 PT products for, e.g. treating neurologic disorders, angiogenesis  
 PT disorders, tumours or rheumatoid arthritis or for wound healing  
 XX  
 PS Claim 20; Fig 1A-C; 86pp; English.  
 XX  
 CC This is a AL-2-long (AL-21) protein. The AL-2 is a novel Eph-related  
 CC tyrosine kinase receptor ligand. AL-2 can be administered to patients in  
 CC whom the nervous system has been damaged by trauma, surgery, stroke,  
 CC ischemia, infection, metabolic disease, nutritional deficiency, or  
 CC malignancy, or toxic agents, to promote the survival or growth of  
 CC neurons. They can be used to treat motoneuron disorders such as  
 CC amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and  
 CC various conditions involving spinal muscular atrophy, or paralysis. AL-2  
 CC can be used to treat human neurodegenerative disorders, such as  
 CC Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating  
 CC diseases such as multiple sclerosis, Huntington's chorea, Down's syndrome,  
 CC nerve deafness, Menier's disease, and other disorders of the cerebellum.  
 CC AL-2 can be used as cognitive enhancer, to enhance learning particularly  
 CC in dementia or trauma, since they can promote axonal outgrowth and  
 CC synaptic plasticity, particularly of hippocampal neurons that express  
 CC AL-2 binding Eph-family receptors and cortical neurons that express  
 CC AL-2. AL-2 can also be used for wound healing, i.e. accelerating  
 CC neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids  
 CC are useful in preparing antibodies that specifically bind to the AL-2  
 CC protein. The antibodies and the AL-2 antagonists are useful in diagnosing  
 CC and treating various neuronal disorders. AL-2 antagonists can be used  
 CC for modulating angiogenesis. They can also be used for the treatment of  
 CC tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),  
 CC myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular  
 CC glaucoma, psoriasis and rheumatoid arthritis.  
 CC  
 CC Sequence 455 AA;  
 SQ  
 Query Match 100.0%; Score 2450; DB 18; Length 455;  
 Best Local Similarity 100.0%; Pred. No. 2, 5e-198;  
 Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGPHSGPGGVRGALLLGLVGLVSGLSLEPYWNSANKRFOAEGGYLYPIGRLDL 60  
 DB 1 MGPHSGPGGVRGALLLGLVGLVSGLSLEPYWNSANKRFOAEGGYLYPIGRLDL 60  
 QY 1 LCPARAPPGHSSSNVEFYKLYVGAQGRCAFPAPNLLTCDDPDDLRFTTFQRY 120  
 DB 1 LCPARAPPGHSSSNVEFYKLYVGAQGRCAFPAPNLLTCDDPDDLRFTTFQRY 120  
 QY 61 LCPARAPPGHSSSNVEFYKLYVGAQGRCAFPAPNLLTCDDPDDLRFTTFQRY 120  
 DB 61 LCPARAPPGHSSSNVEFYKLYVGAQGRCAFPAPNLLTCDDPDDLRFTTFQRY 120  
 QY 121 SPNIWGEFRRSHDYIATSDGTREGLESLOQGVCTTRGKYLRLVQGSFRGAAVPRKP 180  
 DB 121 SPNIWGEFRRSHDYIATSDGTREGLESLOQGVCTTRGKYLRLVQGSFRGAAVPRKP 180  
 QY 121 SPNIWGEFRRSHDYIATSDGTREGLESLOQGVCTTRGKYLRLVQGSFRGAAVPRKP 180  
 DB 121 SPNIWGEFRRSHDYIATSDGTREGLESLOQGVCTTRGKYLRLVQGSFRGAAVPRKP 180  
 QY 161 VSEPMRDRGAHSLERPKENLEGPDTSNATSGAAGPLPPSPMPAVAGAAGLALLL 240  
 DB 161 VSEPMRDRGAHSLERPKENLEGPDTSNATSGAAGPLPPSPMPAVAGAAGLALLL 240  
 QY 161 VSEPMRDRGAHSLERPKENLEGPDTSNATSGAAGPLPPSPMPAVAGAAGLALLL 240  
 DB 161 VSEPMRDRGAHSLERPKENLEGPDTSNATSGAAGPLPPSPMPAVAGAAGLALLL 240  
 QY 241 GVAAGAGMCMRRRAKPPSRRHPGSGFRGSGIGVGGGGMGPRAAFGELGIALRG 300  
 DB 241 GVAAGAGMCMRRRAKPPSRRHPGSGFRGSGIGVGGGGMGPRAAFGELGIALRG 300  
 QY 301 GAADPPCPHYEKSGDYGHPIVVDGPPSPNNIYTSISVLEWPIILHTIQLFFMRSK 360  
 DB 301 GAADPPCPHYEKSGDYGHPIVVDGPPSPNNIYTSISVLEWPIILHTIQLFFMRSK 360

DB 301 GAADPPCPHYEKSGDYGHPIVVDGPPSPNNIYTSISVLEWPIILHTIQLFFMRSK 360  
 QY 361 CSRVTTLFPVOVITTTCTMTSSFTTLNPSMOACRAQGEFRIRRCFPGDRIIGTALF 420  
 DB 361 CSRVTTLFPVOVITTTCTMTSSFTTLNPSMOACRAQGEFRIRRCFPGDRIIGTALF 420  
 QY 421 VLVLILLGRINMQTTLRORASVEAAGQHP 455  
 DB 421 VLVLILLGRINMQTTLRORASVEAAGQHP 455  
 RESULT 2  
 ID AAW33699 standard; Protein; 340 AA.  
 XX AAW33699  
 AC AAW33699;  
 XX  
 DT 30-APR-1998 (first entry)  
 XX  
 DE AL-2-short (AL-2s) protein.  
 XX  
 KW AL-2s; AL-2; AL-2-short; human; treatment; neurological disorder; tumour;  
 KW rheumatoid arthritis; wound healing; paralysis; angiogenesis; leukaemia;  
 KW psoriasis; Alzheimer's disease; epilepsy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..26  
 FT /note= "signal peptide"  
 FT Protein 27..340  
 FT /note= "mature protein"  
 FT Domain 27..219  
 FT /note= "extracellular domain"  
 FT Domain 220..245  
 FT /note= "hydrophobic transmembrane domain"  
 XX  
 PN MO9740153-AL.  
 XX  
 PD 30-OCT-1997.  
 XX  
 PF 17-APR-1997; 97WO-US06345.  
 XX  
 PR 19-APR-1996; 96US-0635130.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Caras IW;  
 XX  
 DR WPI, 1997-535837/49.  
 XX  
 DR N-PSDB; AAV06355.  
 XX  
 PT Human AL-2 neurotrophic factor and related DNA - used to develop  
 PT products for, e.g. treating neurologic disorders, angiogenesis  
 PT disorders, tumours or rheumatoid arthritis or for wound healing  
 XX  
 PS Claim 20; Fig 2A-B; 86pp; English.  
 XX  
 CC This is a AL-2-short (AL-2s) protein. The AL-2 is a novel Eph-related  
 CC tyrosine kinase receptor ligand. AL-2 can be administered to patients in  
 CC whom the nervous system has been damaged by trauma, surgery, stroke,  
 CC ischemia, infection, metabolic disease, nutritional deficiency, or  
 CC malignancy, or toxic agents, to promote the survival or growth of  
 CC neurons. They can be used to treat motoneuron disorders such as  
 CC amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and  
 CC various conditions involving spinal muscular atrophy, or paralysis. AL-2  
 CC can be used to treat human neurodegenerative disorders, such as  
 CC Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating  
 CC diseases such as multiple sclerosis, Huntington's chorea, Down's syndrome,  
 CC nerve deafness, Menier's disease, and other disorders of the cerebellum.  
 CC AL-2 can be used as cognitive enhancer, to enhance learning particularly  
 CC in dementia or trauma, since they can promote axonal outgrowth and  
 CC synaptic plasticity, particularly of hippocampal neurons that express



CC AL-2 binding Eph-family receptors and cortical neurons that express  
 CC AL-2. AL-2 can also be used for wound healing, i.e. accelerating  
 CC neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids  
 CC are useful in preparing antibodies that specifically bind to the AL-2  
 CC protein. The antibodies and the AL-2 antagonists are useful in diagnosing  
 CC and treating various neuronal disorders. AL-2 antagonists can be used  
 CC for modulating angiogenesis. They can also be used for the treatment of  
 CC tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),  
 CC myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular  
 CC glaucoma, psoriasis and rheumatoid arthritis.  
 CC  
 SQ Sequence 340 AA;  
 Query Match 75.1%; Score 1841; DB 18; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-147;  
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGPHSGPGVGVGALLLGLVGLVSGLSLEPVYVNSANKRFOAEGGYLYPQIGDRDL 60  
 DB 1 MGPHSGPGVGVGALLLGLVGLVSGLSLEPVYVNSANKRFOAEGGYLYPQIGDRDL 60  
 QY 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLTCRDPDLRFTIKQRY 120  
 DB 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLTCRDPDLRFTIKQRY 120  
 QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGVCLTRGMKVLRLVGOSPRGAVPRKP 180  
 DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGVCLTRGMKVLRLVGOSPRGAVPRKP 180  
 QY 181 VSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLL 240  
 DB 181 VSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLL 240  
 QY 241 GVAAGAGAMCWRRRRAKPSBSRHPGSGFGRGSLGAGGGGMPREAPGELIALRG 300  
 DB 241 GVAAGAGAMCWRRRRAKPSBSRHPGSGFGRGSLGAGGGGMPREAPGELIALRG 300  
 QY 301 GAADPPFCPHYEKVSGDYGHPIYIVODGPSPNNIYY 338  
 DB 301 GAADPPFCPHYEKVSGDYGHPIYIVODGPSPNNIYY 338  
 RESULT 3  
 AAW31544  
 ID AAW31544 standard; Protein; 340 AA.  
 XX  
 AC AAW31544;  
 XX  
 DT 14-APR-1998 (first entry)  
 XX  
 DE Human cytokine Lerk-8.  
 XX  
 KW Lerk-8; cytokine; human; hek; elk; receptor tyrosine kinase;  
 KW ligand; neurodegenerative disease; wound healing;  
 KW neovascularisation; diagnosis; therapy.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /label= Sig\_peptide  
 FT Protein /label= Mat\_protein  
 FT Domain 28..224  
 FT Domain /note= "extracellular domain"  
 FT Domain 225..251  
 FT Domain /note= "transmembrane domain"  
 FT Domain 252..340  
 FT Modified-site /note= "cytoplasmic domain"  
 FT Modified-site 210..212  
 FT Misc-difference /note= "N-glycosylated"  
 FT /note= "residue 325 is Leu in Lerk-8 variant"

XX  
 PN WO9356919-A2.  
 PD 09-OCT-1997.  
 XX  
 PF 19-MAR-1997; 97WO-US04533.  
 XX  
 PR 21-MAR-1996; 96US-0621146.  
 XX  
 PA (IMMUV) IMMUNEX CORP.  
 XX  
 PI Cerretti DP;  
 XX  
 DR WPI, 1997-503043/46.  
 DR N-PSDB; AAT89519.  
 PT New isolated cytokine, Lerk-8 - binds to the hek and elk receptor  
 PS Tyrosine kinases, used to develop products for diagnosis and therapy  
 XX  
 Claim 3; Page 32-33; 37pp; English.

CC This protein sequence comprises a novel human cytokine designated  
 CC Lerk-8. The amino acid sequence was deduced from a human foetal  
 CC brain cDNA clone (see AAT89519). Lerk-8 binds to the cell surface  
 CC receptors hek and elk, which are members of the eph/elk family of  
 CC receptor tyrosine kinases. Lerk-8 polypeptides, especially soluble  
 CC polypeptides comprising amino acid residues -27 to 142-197 of the  
 CC full-length protein, can be expressed in transformed host cells.  
 CC These polypeptides can be used to purify hek or elk proteins, and  
 CC to purify or identify cells that express hek or elk on the surface.  
 CC Such cells can be used in various in vitro studies or in vivo  
 CC procedures, e.g. neural cells expressing elk can be administered to  
 CC a mammal afflicted with a neurodegenerative disorder. The Lerk-8  
 CC polypeptides can also be used to deliver diagnostic or therapeutic  
 CC agents to these cells (e.g. leukaemia cells). The Lerk-8 DNA and  
 CC polypeptides can also be used to treat disorders mediated by  
 CC defective or insufficient amounts of Lerk-8; to treat disorders  
 CC such as injury to neural tissue or neurologic disease; to promote  
 CC angiogenesis; and for wound healing or stimulating  
 CC neovascularisation of grafted tissues.  
 CC  
 XX

SQ Sequence 340 AA;  
 Query Match 75.1%; Score 1841; DB 18; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-147;  
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MGPHSGPGVGVGALLLGLVGLVSGLSLEPVYVNSANKRFOAEGGYLYPQIGDRDL 60  
 DB 1 MGPHSGPGVGVGALLLGLVGLVSGLSLEPVYVNSANKRFOAEGGYLYPQIGDRDL 60  
 QY 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLTCRDPDLRFTIKQRY 120  
 DB 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLTCRDPDLRFTIKQRY 120  
 QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGVCLTRGMKVLRLVGOSPRGAVPRKP 180  
 DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGVCLTRGMKVLRLVGOSPRGAVPRKP 180  
 QY 181 VSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLL 240  
 DB 181 VSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLL 240  
 QY 241 GVAAGAGAMCWRRRRAKPSBSRHPGSGFGRGSLGAGGGGMPREAPGELIALRG 300  
 DB 241 GVAAGAGAMCWRRRRAKPSBSRHPGSGFGRGSLGAGGGGMPREAPGELIALRG 300  
 QY 301 GAADPPFCPHYEKVSGDYGHPIYIVODGPSPNNIYY 338  
 DB 301 GAADPPFCPHYEKVSGDYGHPIYIVODGPSPNNIYY 338  
 RESULT 4

AAW10637  
ID AAW10637 standard; Protein; 340 AA.  
XX  
AC AAW10637;  
XX  
DT 23-JUN-1997 (first entry)  
XX  
DE NLERK2 ligand for eph-related kinase.  
XX  
KM LERK, ligand for eph-related kinase; ERK, NLERK2;  
XX  
KM receptor protein tyrosine kinase; cell proliferation;  
XX  
KM cell differentiation; cell survival; nerve cell.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..29  
FT /label= Sig\_peptide  
FT Protein 30..340  
FT /label= Mat\_protein  
FT Domain 227..251  
FT /label= Transmembrane\_domain  
FT Modified-site 210  
FT /label= N-glycosylation\_site  
XX  
PN WO9704091-A1.  
XX  
PD 06-FEB-1997.  
XX  
PF 19-JUL-1996; 96WO-AU00460.  
XX  
PR 05-FEB-1996; 96AU-0007890.  
XX  
PR 20-JUL-1995; 95AU-0004263.  
XX  
PR 27-NOV-1995; 95AU-0006847.  
XX  
PR 22-DEC-1995; 95AU-0007299.  
XX  
XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
XX  
XX Nicotiana glauca.  
XX  
PI WPI; 1997-132632/12.  
XX  
DR N-PSDB; AAT60966.  
XX  
XX Nucleic acid mol. encoding ligand for eph-related kinase - useful  
PT for treatment of, pref. neuronal, cells to increase survival,  
PT proliferation and differentiation  
XX  
PS Claim 16; Page 37-39; 71pp; English.  
XX  
XX A novel human ligand for eph-related kinase (LERK) is designated  
CC NLERK2 (AAW10637). It is encoded by a cDNA clone (AAT60966) obt.  
CC from a human foetal brain cDNA library. The novel receptor ligand  
CC can be expressed in transformed host cells and used in methods  
CC for regulating the development, maintenance or regeneration of  
CC different cells (e.g. neurons) and tissues in vitro and in vivo.  
CC Soluble NLERK2 peptides can be used to treat injury, disease or  
CC abnormality in the nervous system, and membrane-bound NLERK2 to  
CC modulate proliferation, different or survival e.g. in grafting  
CC procedures or transplantation. NLERK2 can also be used to raise  
CC antibodies for use in immunotherapy, and to detect anti-NLERK2  
CC antibodies that may occur in some autoimmune diseases.  
XX  
PS Sequence 340 AA;  
XX  
Query Match 75.1%; Score 1841; DB 18; Length 340;  
Best Local Similarity 100.0%; Pred. No. 4,4e-147;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 61 LCPARPFGHSSPNFYKYLYLVGAGQRCRCAFPAPNLLITCDRDLRLFTIKQXY 120  
QY 121 SPNLMGHEFRSHDYIATISDGTREGLESLOQGVCLTRGMKYLRLVQSPRGCAVRRK 180  
DB 121 SPNLMGHEFRSHDYIATISDGTREGLESLOQGVCLTRGMKYLRLVQSPRGCAVRRK 180  
QY 181 VSEMPMERDRAAHLTEPKENLPGDPTSNATSRGAEGPLPPSPMPAVAGAGLALLL 240  
DB 181 VSEMPMERDRAAHLTEPKENLPGDPTSNATSRGAEGPLPPSPMPAVAGAGLALLL 240  
QY 241 GVAGAGGAMCWRRRAPSRPSRHPGSGFGRGSLGLGGGGMGPRABEGELGIALRG 300  
DB 241 GVAGAGGAMCWRRRAPSRPSRHPGSGFGRGSLGLGGGGMGPRABEGELGIALRG 300  
QY 301 GAADPFPCPHYEKVSGDYGHPIYIVQDGPQSPNNIY 338  
DB 301 GAADPFPCPHYEKVSGDYGHPIYIVQDGPQSPNNIY 338

## RESULT 5

AAW17081  
ID AAW17081 standard; Protein; 340 AA.

AAW17081;  
AC

09-AUG-1997 (first entry)  
DT

Eph family ligand Efl-6.  
DE

Efl-6; Eph; Etk, receptor tyrosine kinase; signal transduction;  
XX  
KM ligand; neurological disease.  
XX

Homo sapiens.  
OS

XX  
FH

Key Location/Qualifiers  
FT Peptide 1..24  
FT /label= Sig\_peptide  
FT Protein 25..340  
FT /label= Mat\_protein  
FT Domain 225..249  
FT /label= Transmembrane\_domain  
FT Misc-difference 166  
FT /label= Gln, Arg

WO9715667-A1.  
XX  
PD 01-MAY-1997.  
XX  
PF 25-OCT-1996; 96WO-US17201.  
XX  
PR 25-OCT-1995; 95US-0007015.  
XX  
PA (REGG-) REGENERON PHARM INC.  
XX  
PI Davis S, Gale NW, Yancopoulos GD;  
XX  
DR WPI; 1997-259021/23.  
XX  
DR N-PSDB; AAT69808.  
XX  
PT New nucleic acid encoding Efl-6 ligand protein - used for promoting  
PT growth and proliferation of neuronal cells and in drug screening  
XX  
PS Claim 2; Fig 1; 36pp; English.

XX A novel ligand (AAW17081), designated Efl-6 (or Eph transmembrane  
CC tyrosine kinase family ligand 6), binds to the Etk, Ntk/Cek5,  
CC Hek2/Sek4, Htk and Sek1 receptors on cells. Its amino acid  
CC sequence was deduced from a human frontal cortex cDNA clone  
CC (AAT69808). Recombinant Efl-6, truncated soluble polypeptides  
CC comprising the extracellular domain of Efl-6, and Efl-6  
CC ligandodies comprising soluble Efl-6 and the FC portion of IgG can  
CC be expressed in host cells. These can be used to support neuronal

CC and other Eph receptor-bearing cell populations for treatment of  
CC neurological disorders, in drug screening and to raise diagnostic  
CC antibodies.

XX Sequence 340 AA;

Query Match 74.9%; Score 1835; DB 18; Length 340;  
Best Local Similarity 99.7%; Pred. No. 1.4e-146;  
Matches 337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGPPHSGGVRVGLLLGLVLSGLSEPVYVNSANKRFQAEQGVLYTQIGDRDL 60  
DB 1 MGPPHSGGVRVGLLLGLVLSGLSEPVYVNSANKRFQAEQGVLYTQIGDRDL 60  
QY 61 LCPRRAPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRPLDLRFTIKFOEY 120  
DB 61 LCPRRAPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRPLDLRFTIKFOEY 120  
QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
QY 241 GVAAGAGAMCWRRRRAKSPESRHPGSGFRGSLGLGGGGGMPREAPBGLIALRG 300  
DB 241 GVAAGAGAMCWRRRRAKSPESRHPGSGFRGSLGLGGGGGMPREAPBGLIALRG 300  
QY 301 GAADPPFCPHYEKVGSDYGHPIYIVDGPPOSPPNITY 338  
DB 301 GAADPPFCPHYEKVGSDYGHPIYIVDGPPOSPPNITY 338

# RESULT 6

AAW46615  
ID AAW46615 standard; Protein; 340 AA.

XX AAW46615;

XX 06-JUL-1998 (first entry)

XX Human transmembrane ligand Elk-L3.

KM Elk-L3; Elk-related receptor tyrosine kinase; transmembrane ligand;  
KM human; signal transduction; axonogenesis; nerve cell; neurone;  
KM Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KM demyelination; multiple sclerosis; amyotrophic lateral sclerosis;  
KM nervous system infection; Wernicke's disease; trauma; ischemia;  
KM stroke; nutritional polyneuropathy; progressive supranuclear palsy;  
KM Shy Drager's syndrome; multistem degeneration;  
KM olivoponto cerebellar atrophy; peripheral nerve damage.

XX Homo sapiens.

XX Key location/Qualifiers

XX Misc-difference 166 /Label= Gln, Arg

XX FT Domain 225..249 /note= "transmembrane domain"

XX MO9801548-A1.

XX 15-JAN-1998.

XX 04-JUL-1997; 97WO-CA00473.

XX 05-JUL-1996; 96US-0021272.

XX (MOUNT ) MOUNT SINAI HOSPITAL CORP.

XX Holland S, Mdamalu G, Pawson T;

XX WPI; 1998-101047/09.  
DR N-PSDB; AAV16097.

XX Modulating transmembrane ligand for an Elk-related receptor tyrosine

PT kinase - by formation of a complex between an oligomerized

XX Elk-related receptor tyrosine kinase and a transmembrane ligand

PS Disclosure; Fig 5A; 40pp; English.

CC This polypeptide comprises human Elk-L3, a transmembrane ligand of  
CC Elk-related receptor tyrosine kinase (ERK). A novel method of  
CC modulating the biological activity of, or for affecting a pathway  
CC regulated by, a transmembrane ligand for an ERK in a cell  
CC expressing the transmembrane ligand comprises forming a complex  
CC between a purified and isolated oligomerized ERK, or an isoform  
CC or an extracellular domain of the ERK, and the transmembrane  
CC ligand expressed on the cell. The complex can also be used for  
CC evaluating a substance for its ability to modulate the biological  
CC activity of a transmembrane ligand for an ERK, and to identify  
CC substances that affect or modulate a pathway regulated by a ERK.  
CC A purified and isolated oligomerized ERK can be used in the  
CC preparation of a medicament for modulating neuronal development or  
CC regeneration in a subject, or in a medicament for modulating  
CC axonogenesis in a subject (all claimed). The substances identified  
CC by the methods can be used to modulate axonogenesis, nerve cell  
CC interactions and regeneration, to treat diseases and conditions  
CC involving trauma and injury to the nervous system, such as  
CC Alzheimer's disease, Parkinson's disease, Huntington's disease,  
CC demyelinating diseases, such as multiple sclerosis, amyotrophic  
CC lateral sclerosis, bacterial and viral infections of the nervous  
CC system, deficiency diseases, such as Wernicke's disease and  
CC nutritional polyneuropathy, progressive supranuclear palsy,  
CC Shy Drager's syndrome, multistem degeneration and Olivoponto  
CC cerebellar atrophy, peripheral nerve damage, trauma, and ischaemia  
CC resulting from stroke.

XX Sequence 340 AA;

Query Match 74.6%; Score 1827; DB 19; Length 340;  
Best Local Similarity 99.4%; Pred. No. 6.7e-146;  
Matches 336; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGPPHSGGVRVGLLLGLVLSGLSEPVYVNSANKRFQAEQGVLYTQIGDRDL 60  
DB 1 MGPPHSGGVRVGLLLGLVLSGLSEPVYVNSANKRFQAEQGVLYTQIGDRDL 60  
QY 61 LCPRRAPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRPLDLRFTIKFOEY 120  
DB 61 LCPRRAPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRPLDLRFTIKFOEY 120  
QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
QY 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYITATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
QY 181 VSEMPERDRGAHSLPEKENVLPDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLL 240  
DB 181 VSEMPERDRGAHSLPEKENVLPDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLL 240  
QY 241 GVAAGAGAMCWRRRRAKSPESRHPGSGFRGSLGLGGGGGMPREAPBGLIALRG 300  
DB 241 GVAAGAGAMCWRRRRAKSPESRHPGSGFRGSLGLGGGGGMPREAPBGLIALRG 300  
QY 301 GAADPPFCPHYEKVGSDYGHPIYIVDGPPOSPPNITY 338  
DB 301 GAADPPFCPHYEKVGSDYGHPIYIVDGPPOSPPNITY 338

# RESULT 7

AAW00287  
ID AAW00287 standard; Protein; 334 AA.

XX AAW00287;

XX	19-JAN-1997	(first entry)
XX	Mouse Eph receptor ligand ELF-2.	
XX	Eph receptor; ligand; ELF-2; tyrosine kinase; signal transduction;	
XX	organogenesis; oncogenesis; tumour; neurological disorder;	
XX	diagnosis; gene therapy.	
XX	Mus sp.	
XX	Key	
XX	Location/Qualifiers	
XX	Domain	
XX	1..224	
XX	/label= Extracellular_domain	
XX	Binding-site	
XX	31..155	
XX	/label= Receptor_binding_site	
XX	Domain	
XX	226..251	
XX	/label= Transmembrane_domain	
XX	MO9626958-A2.	
XX	06-SEP-1996.	
XX	23-FEB-1996; 96WO-US02673.	
XX	27-FEB-1995; 95US-0395415.	
XX	(HARD ) HARVARD COLLEGE.	
XX	Bergemann AD, Flanagan JG;	
XX	WPI; 1996-43391/43.	
XX	N-PSDB; AAT40230.	
XX	Eph receptor ligand, ELF-2, and DNA encoding it - used to treat or	
XX	prevent neurological diseases, and to modulate binding of ELF-2 to	
XX	Eph receptor, e.g. to prevent or treat tumour formation.	
XX	Claim 6; Fig 1A-B; 50pp; English.	
XX	Mouse Eph receptor ligand ELF-2 (AA00287) is strongly expressed in	
XX	the anterior hindbrain and newly-forming somites of embryos at the	
XX	early organogenesis stage of development. It is important in	
XX	cellular communication during pattern formation. Its amino acid	
XX	sequence was deduced from a cDNA clone (AAT40230) isolated from a	
XX	newborn mouse brain cDNA library. The ELF-2 ligand can be used	
XX	to alter neurological development, oncogenesis and growth	
XX	regulation, to modulate binding of ELF-2 to the Eph receptor, and	
XX	in diagnostic assays.	
XX	Sequence 334 AA;	
XX	Query Match 25.7%; Score 628.5; DB 17; Length 334;	
XX	Best Local Similarity 41.8%; Pred. No. 9,5e-45;	
XX	Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5	
XX	14 GALLITGLVIGVLSGLSEPIYNNNSAKRRQAGGYLYPQIGDRLLDLCPRARPPGHSS 73	
XX	15 GLIWLVCRTAISRSITVEPIYNNSSKFTLPGQGLVLPQIGDKDITCPKY--DSKITV 71	
XX	74 PNYEYFYLYLVGAQGRCEAPPAENLLITCDRFDLIDRFITIKFOEYSPNLMGHEPFSHH 133	
XX	72 GQYEYKYVYVWDKQOARCTIKKENPILNLCARPDQDVYFTIKFOEFPNNLMGHEPFSKH 131	
XX	134 DYVLIANSQGRRELSLOGVCLTGGMVTLARVQG--SPRGAVPRRVSVEPMER-DR 190	
XX	132 DYVLIISNSQGLBGLNQEGGVCQITAMKILMKVGGDASSAGSARNHGPTRPELEAGTN 191	
XX	191 GAASLBEFGKENTPGDTSNATSGAEGPLPPSPPAVAGAAGLALLLVAGAAGAMC 250	
XX	192 GRSSSTSPFYKPNFGSSTDGNASGHSGNNLLGSEVALFAGIASGCIITVILITLVLL 251	
XX	251 WRRRAAPSEERNAHGPDSFGGSGSLGGGGGGMGPENAPSGSLGIALGGGAADPPCPH 310	

Db	252	KYRRRHRSQHCHTTTSLSLATLTPKGGNN----	NGSEPSDVIILPLR--	TALDSVCFPH	304
Oy	311	YKVSQDYGHPVYIVQDGPSPSPNITY	338		
Db	305	YKVSQDYGHPVYIVQDGPSPSPNITY	332		
RESULT 8					
AAR92742					
ID	AAR92742	standard; Protein; 336 AA.			
XX					
AC	AAR92742;				
XX					
DT	21-MAY-1996	(first entry)			
DE	Murine hepatoma transmembrane kinase receptor ligand.				
XX					
KW	Hepatoma transmembrane kinase; Htk; receptor; ligand;				
XX	tyrosine kinase; neurodegenerative disease.				
OS	Mus musculus.				
XX					
EN	W09602645-A2.				
XX					
PD	01-FEB-1996.				
XX					
PF	14-JUL-1995; 95WO-US08812.				
XX					
FR	20-JUL-1994; 94US-0277722.				
XX					
PA	(GETH ) GENENTECH INC.				
XX					
PI	Bennett BD, Matthews W;				
XX					
DR	WPI; 1996-105907/11.				
XX	N-PSDB; AAT16470.				
PT	Ligand for the hepatoma trans-membrane kinase receptor - useful for				
PT	stimulating and inhibiting cells carrying the receptor, e.g. for				
PT	treating neuro-degenerative disease				
XX					
PS	Claim 5; Fig 1(A-D); 88pp; English.				
XX					
CC	Mouse (AAT16470) and human (AAT16471) Htk ligand which bind to, and				
CC	activate, the Htk receptor, have been identified in a variety of				
CC	tissues using a soluble Htk-Fc fusion protein.				
CC	The predicted mol.wt. of the murine Htk ligand protein following a				
CC	signal peptide cleavage is 34 kD with an estimated pI of 8.9.				
CC	The murine and human ligands show 96% homology at the amino acid				
CC	level.				
CC	The DNA is used to produce recombinant ligands; for tissue-				
CC	specific typing (partic. as a marker for breast cancer) and as a				
CC	marker for human chromosome 13. The ligands (partic. in soluble				
CC	form) are used to activate the tyrosine kinase domain of the				
CC	Htk receptor, i.e. to stimulate or inhibit growth, differentiation,				
CC	and/or activation of cells contg. the receptor, e.g. treatment				
CC	of neurodegenerative diseases, since they are strongly expressed				
CC	in the cerebral cortex, hippocampus, striatum and cerebellum.				
CC	The ligands are also useful as a control or standard in assays,				
CC	for generation of antibodies, as a mol. wt. marker, for growth				
CC	in vitro of Htk-receptor positive cells, as research agent,				
CC	in screening, etc.				
XX					
SQ	Sequence 336 AA;				
Query Match	25.7%; Score 628.5; DB 17; Length 336;				
Best Local Similarity	41.8%; Pred. No. 9.6e-45;				
Matches 137; Conservative	49; Mismatches 129; Indels 13; Gaps				
Oy	14	GALLITGLVIGVSGSLTPPYWNSANKRFQAGGVLTLPQIGDRLLDLCPPARPPGPHSS	73		
Db	17	GLLMLVCLRTAISRSIVLEPIYWNSSNFKFLPQGGVLVLPQIGDKDIDICPKV--	DSKTV	73	



XX The present sequence is the human cytokine  $\text{elk-L}$  ligand ( $\text{elk-L}$ ),  
 CC which binds a member of the tyrosine kinase receptor family.  $\text{elk-L}$   
 CC exhibits neurotrophic and neuroprotective properties, and has a  
 CC calculated mol. wt. 35180 daltons and a pI of 9.006. The  $\text{elk-L}$   
 CC cDNA, isolated from a human placental cDNA library, can be  
 CC radiolabeled and used as a probe for isolating other mammalian  
 CC  $\text{elk-L}$  cDNA.  $\text{elk-L}$  can be used to treat neural tissue disorders,  
 CC partic. excitotoxicity associated injuries or disorders, and as a  
 CC neural culture reagent, while immunogenic fragments of  $\text{elk-L}$  can be  
 CC used to generate specific anti- $\text{elk-L}$  antibodies.

XX Sequence 346 AA;

Query Match 25.4%; Score 623; DB 17; Length 346;

Best Local Similarity 39.2%; Pred. No. 2.9e-44; Mismatches 116; Indels 58; Gaps 9;

Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGVAVGALLLLGLVLSGL-----SLRPVYNSANKRFOAGGVLYLPQIGRLDLL 61  
 DB 4 PGQRLGKMLVAMVWMLCRLATPLAKNLEPVSWSLNPKFLSGKGLVYIKIGDKLDII 63

QY 62 CPRARPPGPHSSPNYEFYKLYLVGAQGRCEAPAPNLLITCDRDLIRFTIKFOEYS 121  
 DB 64 CPRAEAGRP-----YEFYKLYLVPEQAAACSTVLDPNVLTGCRPEOEIRFTIKFOEYS 118

QY 122 PNLWGHFRSHHDYIITATSDGTREGLSELOGVCLTRGKMLLRVQSPGAVPRKPY 181  
 DB 119 PNIWGLDFPKKHHDYITSTNGSLSELENREGVCRTTKIMKYGQDPNATPPQLTT 178

QY 182 SEMPERDRGAHSLF-PGKENTPGDPTSNATSGAEGPLPPSPMAVAGAGLA---- 236  
 DB 179 SRPSKEADNTVMATQAPGSRGSLGDSGKHEVTNQEKSFP-----GASGSSGSDPD 231

QY 237 -----LLILGVAGGA-----MCKRRRAKSESRRHPGSGFSGSLGI 277  
 DB 232 GFENSKVALFAVAGACVITFLIIITVLLIKRKRRKHQQ-----RAALSLI 282

QY 278 -----GGGCGMGPREAPGELGIALRGGAADPPCPHYEKVSGDYGHPVYIVODGPPQSP 333  
 DB 283 STLASPKGSGTAGTEPSDIIIPLR---TTENNVCPIHEKXVSGDYGHPVYIVQEMPPQSP 339

QY 334 PNITY 338  
 DB 340 ANITY 344

# RESULT 11

AAW36055 ID AAW36055 standard; Protein; 346 AA.

XX AC AAW36055;

XX DT 06-MAR-1998 (first entry)

XX DE Human  $\text{elk-L}$  protein.

XX KM Human;  $\text{elk-L}$ ; cytokine; ligand; tyrosine kinase receptor; fusion protein;

XX KM extracellular domain; immunoglobulin; neurological disease.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..24 /note= "signal peptide"

FT Protein 25..346 /note= "mature protein"

FT Domain 25..237 /note= "extracellular domain; this region is used to generate a fusion protein with the Fc portion of the human immunoglobulin G1"

FT Domain 238..265 /note= "transmembrane domain"

FT Domain 265..346 /note= "intracellular domain"  
 FT Modified-site 139..141 /note= "Asn is N-glycosylated"  
 FT Cleavage-site 265..267 /note= "KEX2 protease cleavage site"  
 FT Cleavage-site 267..268 /note= "KEX2 protease cleavage site"  
 FT Cleavage-site 270..271 /note= "KEX2 protease cleavage site"

XX US5670625-A.

XX 23-SEP-1997.

XX 02-JUN-1995; 95US-0460741.

XX 15-MAR-1994; 94US-0213403.

XX 13-NOV-1992; 92US-0977693.

XX 02-JUN-1995; 95US-0460741.

XX (IMMUNEX CORP.

XX Baum PR, Beckmann MP, Lyman S;

XX WPI; 1997-479524/44.

XX N-PSDB; AAT97976.

XX Soluble fusion proteins of human  $\text{elk-L}$  ligand and Fc immunoglobulin  
 fragment - and their dimers and oligomers, useful as  
 neuro-protectants and neurotrophic agents

XX Claim 1; Columns 27-30; 18pp; English.

XX This is the amino acid sequence of the human  $\text{elk-L}$  protein, a new  
 CC cytokine that is the ligand for the  $\text{elk}$  tyrosine kinase receptor. The  
 CC extracellular domain of the protein (amino acids 1-213) is used to  
 CC generate a fusion protein comprising the Fc polypeptide of the human  
 CC immunoglobulin G1 (extending from the hinge region to the C-terminus).  
 CC The fusion protein (which has the same activities as the natural  $\text{elk-L}$   
 CC protein) has neuroprotective and neurotrophic activity so is potentially  
 CC useful for treating a wide range of neurological diseases.

XX Sequence 346 AA;

Query Match 25.4%; Score 623; DB 18; Length 346;

Best Local Similarity 39.2%; Pred. No. 2.9e-44; Mismatches 116; Indels 58; Gaps 9;

Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGVAVGALLLLGLVLSGL-----SLRPVYNSANKRFOAGGVLYLPQIGRLDLL 61  
 DB 4 PGQRLGKMLVAMVWMLCRLATPLAKNLEPVSWSLNPKFLSGKGLVYIKIGDKLDII 63

QY 62 CPRARPPGPHSSPNYEFYKLYLVGAQGRCEAPAPNLLITCDRDLIRFTIKFOEYS 121  
 DB 64 CPRAEAGRP-----YEFYKLYLVPEQAAACSTVLDPNVLTGCRPEOEIRFTIKFOEYS 118

QY 122 PNLWGHFRSHHDYIITATSDGTREGLSELOGVCLTRGKMLLRVQSPGAVPRKPY 181  
 DB 119 PNIWGLDFPKKHHDYITSTNGSLSELENREGVCRTTKIMKYGQDPNATPPQLTT 178

QY 182 SEMPERDRGAHSLF-PGKENTPGDPTSNATSGAEGPLPPSPMAVAGAGLA---- 236  
 DB 179 SRPSKEADNTVMATQAPGSRGSLGDSGKHEVTNQEKSFP-----GASGSSGSDPD 231

QY 237 -----LLILGVAGGA-----MCKRRRAKSESRRHPGSGFSGSLGI 277  
 DB 232 GFENSKVALFAVAGACVITFLIIITVLLIKRKRRKHQQ-----RAALSLI 282

QY 278 -----GGGCGMGPREAPGELGIALRGGAADPPCPHYEKVSGDYGHPVYIVODGPPQSP 333  
 DB 283 STLASPKGSGTAGTEPSDIIIPLR---TTENNVCPIHEKXVSGDYGHPVYIVQEMPPQSP 339

QY 334 PNITYY 338  
 DB 340 ANITYY 344

## RESULT 12

ID AAW19249 standard; Protein; 346 AA.

AAW19249;

DT 18-AUG-1997 (first entry)

DE Human elk ligand protien.

KW Human; elk; ligand; elk-L; cytokine; testing; measuring;  
 purification; neuroprotection; treatment; diabetic; hereditary;  
 nutritional; neuropathy; neurodegenerative disease;  
 tissue culture.

KM Homo sapiens.

OS Location/Qualifiers

FT Peptide 1..24 /label= sig\_peptide

FT Peptide 25..346 /label= mat\_peptide

FT US5627267-A.

PD 06-MAY-1997.

PF 13-NOV-1992; 92US-0977693.

PR 15-MAR-1994; 94US-0213403.

PR 13-NOV-1992; 92US-0977693.

PR 01-JUN-1995; 95US-0458077.

XX (IMMV ) IMMUNEX CORP.

PI Baum PR, Beckmann MP, Lyman S;

DR WPI; 1997-271366/24.

DR N-PSDB; AAT69766.

FT Human elk ligand protein - for diagnostic or therapeutic use, e.g.  
 as neuro-protective agent

PS Claim 1; Columns 29-32; 18pp; English.

CC The present sequence is a human elk ligand (elk-L) protein,  
 which binds elk, has a calculated molecular weight of 35180 and an  
 isoelectric point of 9.006. Elk-L is a cytokine that can be used to  
 test cells for elk expression, measure the biological activity of  
 elk, purify elk by affinity chromatography and as a neuroprotective  
 agent to treat diabetic, hereditary and nutritional neuropathies  
 CC and neurodegenerative diseases. It may also be added to tissue  
 CC cultures to prolong the life of neurons. The elk-L cDNA was  
 CC isolated from a human placental cDNA library, and is present as a  
 CC cDNA insert in the recombinant vector deposited in strain  
 CC ATCC 69085.

XX Sequence 346 AA;

Query Match 25.4%; Score 623; DB 18; Length 346;

Best Local Similarity 39.2%; Pred. No. 2.9e-44;

Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGCVRGALLLLGVTLGVGL-----SLPEYVNSANKRFOAEGGYLYPIQIDRLDL 61

DB 4 PQGRMLGKLVAMVVALRLATPLAKNIEFVNSINPKFISGKGVLYPIKIGKLDII 63

QY 62 CPRARPPGHSSPNVEFYKLIVGGAQGRCAAPPANILLTCRPPDLLRFTIKFOEYS 121

DB 64 CPRABGRP-----VEYYKLVLPPEQAAACSTVLDPNVLVTCNRPDEGIRFTIKFOEYS 118  
 QY 122 PNLMGHEFRSHHDYIILATSDGTRBGLSELQGVCLTRGMKYLRLRGQSPRGAVPRKRV 181  
 DB 119 PNYMGLEFKKHHDYITTSNSGLELENREGGVCKRTMTKIMKVGODPNNAVTPQLTT 178  
 QY 182 SEMPMERDGAASLE-PEGKENVGDPSTSNATSRGACGLPPEPSMPAVGAAGLA---- 236  
 DB 179 SRPSKEADNTVKATQAPSGRSLGSDGKHETVNOEKSQ-----GASGSGSDPD 231  
 QY 237 -----LLILGVAGAGA-----MCKRRRAKPSSESRHPGSGSRGSGSL 277  
 DB 232 GFNSKVALFAVAGACVIFLLIIFLVTLKLRGRKHTQO-----RAALSL 282  
 QY 278 ----GGGGWGPPEAPSEPELIGLALRGGAADPFCEPKYKSGDYGPVYIVODGPPQSP 333  
 DB 283 STLASPKGSGTACTPSPDIIPLR---TTENNYCHYKVGSDIGHPIYIVQEMPPQSP 339

QY 334 PNITYY 338  
 DB 340 ANITYY 344

## RESULT 13

ID AAW4323 standard; Protein; 346 AA.

AAW4323;

DT 27-MAY-1998 (first entry)

DE Human elk-L.

KW Human; elk-L; elk ligand; cytokine; antibody; cell surface receptor;  
 tyrosine kinase receptor; neural disease; trophic mechanism.

KM Homo sapiens.

OS Location/Qualifiers

FT Peptide 1..24 /label= signal

FT Peptide 25..346 /label= elk-L

FT US5728813-A.

PD 17-MAR-1998.

PF 12-NOV-1996; 96US-0747240.

PR 15-MAR-1994; 94US-0213403.

PR 13-NOV-1992; 92US-0977693.

PR 02-JUN-1995; 95US-0460741.

PR 12-NOV-1996; 96US-0747240.

XX (IMMV ) IMMUNEX CORP.

PI Baum PR, Beckmann MP, Lyman S;

DR WPI; 1998-206621/18.

DR N-PSDB; AAV15226.

FT Antibodies to elk ligand polypeptides - bind to tyrosine kinase  
 PT receptors, useful for treating neural disease

PS Claim 1; Column 31-32; 17pp; English.

CC The present sequence represents human elk-L (elk ligand). The present  
 CC invention describes antibodies to elk-L polypeptides. The elk-L protein  
 CC is known to be neuroprotective exhibiting its effect through a trophic  
 CC mechanism. Examples of diseases that may be treated with elk-L include  
 CC neuropathies e.g. diabetic, hereditary and nutritional neuropathies,



CC neurodegenerative diseases and other diseases characterised by loss of  
 CC function or degeneration of neurons.

XX Sequence 346 AA;

Query Match 25.4%; Score 623; DB 19; Length 346;

Best Local Similarity 39.2%; Pred. No. 2.9e-44;

Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGRVAVALLLVGLVSL-----SLEPYMNSANKRFOAGGVLYLPQIGDLDL 61  
 DB 4 PGGRVAVALLLVGLVSL-----SLEPYMNSANKRFOAGGVLYLPQIGDLDL 63  
 QY 62 CPRARPPGPHSSPYEYKYLIVGAGRCEAPAPNLLTCDRPPDLRFETIKFOEYS 121  
 DB 64 CPRARPPGPHSSPYEYKYLIVGAGRCEAPAPNLLTCDRPPDLRFETIKFOEYS 118  
 QY 122 PNMGHEFRSHHDYIITATSDGTRREGLESIQGVCLTRGMKVLIRVGQSPRGAVPRKRV 181  
 DB 119 PNMGHEFRSHHDYIITATSDGTRREGLESIQGVCLTRGMKVLIRVGQSPRGAVPRKRV 178  
 QY 182 SEMMERDRGAHSLR-PGKENLPDPTSNATSGAEGPLPPSPMVAAGAAGLA----- 236  
 DB 179 SRPEKEDNTVKTATQAPSGRSLGSDGKHETVNOEKSGP-----GASGSSGDD 231  
 QY 237 -----LLLVAGAGGA-----NCMRRRAKPSERHPPGSGFGRGSLGL 277  
 DB 232 GFNSKVALFAAVAGCVIFLLIIFLVTLMLKLRKHRTQO-----RAAALSL 282  
 QY 278 ----GGGGGPRFAEPEGELGIALRGGAADPPFCPHYEKVSGDYGHPIVTVODGPPSP 333  
 DB 283 STLASPKGSGTAGRPEDIIIPLR--TTENNCPHYEKVSGDYGHPIVTVODGPPSP 339

QY 334 PNITY 338  
 DB 340 ANITY 344

#### RESULT 14

AA082606 standard; Protein; 346 AA.

XX AAR82606;  
 AC AAR82606;  
 DT 16-MAY-1996 (first entry)  
 XX Eph transmembrane tyrosine kinase family ligand, Efl-3.  
 DE Eph transmembrane tyrosine kinase family ligand, Efl-3.  
 KW Efl-2; EHK1-L; Eph transmembrane tyrosine kinase family ligand;  
 KW neurological disorder; identification; diagnosis.  
 XX Homo sapiens.

XX Key Location/Qualifiers  
 FH Peptide 1..29  
 FT /label= signal\_peptide  
 FT Domain 238..263  
 FT /label= transmembrane\_domain  
 FT Misc-difference 159  
 FT /note= "borders main conserved regions"

XX W09527060-A2.

XX 12-OCT-1995.

XX 04-APR-1995; 95WO-US04208.

XX 21-OCT-1994; 94US-0327423.

XX 04-APR-1994; 94US-0222075.

XX 12-APR-1994; 94US-0229402.

XX 01-SEP-1994; 94US-0299567.

XX (REG- ) REGENERON PHARM INC.

XX Aldrich TH, Davis S, Gale N, Goldfarb M, Maisompierre PC;  
 PI Yancopoulos GD;  
 XX WPI, 1995-358635/46.

XX Ligands which bind Eph family receptors - used in the diagnosis of  
 PT neurological disorders

XX Disclosure; Fig 2; 58pp; English.

XX Efl-3 (also known as EHK-1) is an Eph transmembrane tyrosine kinase  
 CC family ligand. It has homology with B61 (Efl-1) and Efl-2. Efl-3  
 CC appears to comprise a conventional transmembrane protein with a  
 CC cytoplasmic domain. Efl-3 is useful for identifying other ligands  
 CC for Btk-1, -2, -3, Bck and Elk receptors. The ligands are useful in  
 CC promoting a differential function and/or influencing the phenotype,  
 CC such as growth and/or proliferation, of receptor bearing cells. They  
 CC may be used in the diagnosis, and treatment of neurological disorders.

XX Sequence 346 AA;

Query Match 25.4%; Score 622; DB 16; Length 346;

Best Local Similarity 40.5%; Pred. No. 3.5e-44; Mismatches 105; Indels 52; Gaps 8;

Matches 137; Conservative 44; Mismatches 105; Indels 52; Gaps 8;

QY 29 SLEPYMNSANKRFOAGGVLYLPQIGDLDLCPRARPPGPHSSPYEYKYLIVGAG 88

DB 31 NLEPYMNSANKRFOAGGVLYLPQIGDLDLCPRARPPGPHSSPYEYKYLIVGAG 85

QY 89 GRRCAPAPNLLTCDRPPDLRFETIKFOEYSBNMGHEFRSHHDYIITATSDGTR 148

DB 86 AAGCTVLDPNVLTQCNPEOIRFTIKFOEYSBNMGHEFRSHHDYIITATSDGTR 145

QY 149 ESLQGVCLTRGMKVLIRVGQSPRGAVPRKRVSEMERDRGAHSLR-PGKENLPDP 207

DB 146 ENRBGVCRTRTMKIMKVGDDPNAVTPREGTLTSRPSKEDNTVKTATQAPSGRSLGDS 205

QY 208 TSNATSGAEGPLPPSPMVAAGAAGLA-----LLLVAGAGGA----- 248

DB 206 DQKHETVNOEKSGP-----GASGSSGDDPGFNSKVALFAAVAGCVIFLLIIFL 258

QY 249 ----NCMRRRAKPSERHPPGSGFGRGSLGL-----GGGGGPRFAEPEGELGIALRG 300

DB 259 TVLLKLRKHRTQO-----RAAALSLTASPKGSGTAGRPEDIIIPLR-- 307

QY 301 GAADPPFCPHYEKVSGDYGHPIVTVODGPPSPPNITY 338

DB 308 -TTENNCPHYEKVSGDYGHPIVTVODGPPSPPNITY 344

#### RESULT 15

AA000288 standard; Protein; 331 AA.

XX AA000288;

XX 19-JAN-1997 (first entry)

XX Chicken Eph receptor ligand ELF-2.

XX Eph receptor; ligand; ELF-2; tyrosine kinase; signal transduction;  
 KW organogenesis; oncogenesis; tumour; neurological disorder;  
 KW diagnosis; gene therapy.

XX Gallus sp.

XX W09626958-A2.

XX 06-SEP-1996.

XX 23-FEB-1996; 96WO-US02673.

XX

PR 27-FEB-1995; 95UG-0395415.  
 XX (HARD ) HARVARD COLLEGE.  
 XX Bergemann AD, Flanagan JG;  
 XX WPI, 1996-433391/43.  
 DR N-PSDB; AAT40231.  
 XX  
 XX Eph receptor ligand, ERF-2, and DNA encoding it - used to treat or  
 PT prevent neurological diseases, and to modulate binding of ERF-2 to  
 PT Eph receptor, e.g. to prevent or treat tumour formation.  
 XX  
 PS Claim 6; Fig 2A-D; 50pp; English.  
 XX  
 XX Chicken Eph receptor ligand ERF-2 (AA000288) is a ligand important  
 CC in cellular communication during pattern formation. It is strongly  
 CC expressed in the anterior hindbrain and newly-forming somites of  
 CC embryos at the early organogenesis stage of development. Its amino  
 CC acid sequence was deduced from a cDNA clone (AAT40231) isolated from  
 CC a chicken cDNA library. The ERF-2 ligands can be used to alter  
 CC neurological development, oncogenesis and growth regulation, to  
 CC modulate binding of ERF-2 to the Eph receptor, and in diagnostic  
 CC assays.  
 CC  
 XX Sequence 331 AA;  
 SQ  
 Query Match 25.4%; Score 621.5; DB 17; Length 331;  
 Best Local Similarity 40.2%; Pred. No. 3,7e-44;  
 Matches 133; Conservative 53; Mismatches 124; Indels 21; Gaps 6;  
 QY 14 GALLILGVLGSLGSLSEPVYVNSANKRFOAGGVLYVPOIGDRLDLCPRARPPGHSS 73  
 DB 14 GALLMVFMTALAKSIVLEPIYVNSNPKFLPGQGLVLPQIGDKLDIICPKV--DSKTA 70  
 QY 74 PNYEFKYLIVGAQGRRCCEAPPAENLLTCDRDLDFETIKFOEYSPNLMGHEFRSH 133  
 DB 71 GLYEVYKVMVDKQADSCAIRKONTPLNCAKPDQVKTIFIKFOEYSPNLMGLEFQNK 130  
 QY 134 DYIITATSDGTREGLESLOGGVCLTRGKVTLLRVQSPRGGAIVRKVSEMPMER----- 188  
 DB 131 DYVISTNGSLGELNMGEGVCCQTKMKILMKVGOQDPSAGLPR---SSDPTRRPQEA 187  
 QY 189 -DRGAHSLPEKENVLPDPTSNATSGAEGLPPSPMPAVAGAALALLLVAGAGG 247  
 DB 188 GTNGSSSTTSFVYKDHSSSTDG--SKAGHSILGSEVALAGIASGCIITVITLV 245  
 QY 248 AMCWERRAKPSESHPGSGFGRGSLGLGGGGMGRRAEPGELGIALRGGAADPPF 307  
 DB 246 LLLKRRRRHKSPQHTTLLSLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVF 298  
 QY 308 CPHEKVSQDYGHPYIVODGPPQSPNNIY 338  
 DB 299 CPHEKVSQDYGHPYIVQEMPQSPANNIY 329  
 RESULT 16  
 AA06337  
 ID AA06337 standard; Protein, 333 AA.  
 XX  
 AC AA06337;  
 XX  
 DT 17-JAN-1997 (first entry)  
 XX  
 DE Full length ligand for receptor-type tyrosine kinase protein.  
 XX  
 KM Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;  
 KM human.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 PH Key Location/Qualifiers  
 FT Peptide 1.25

FT Protein /note= "signal peptide"  
 FT 26..333  
 XX  
 XX /note= "receptor-type tyrosine kinase binding protein"  
 PN JP08188596-A.  
 XX  
 PD 23-JUL-1996.  
 XX  
 PF 13-JAN-1995; 95JP-0003677.  
 XX  
 PR 09-NOV-1994; 94JP-0275411.  
 PR 19-OCT-1994; 94JP-0253848.  
 XX  
 XX (ASAH ) ASAH KASEI KOGYO KK.  
 PA  
 DR WPI; 1996-388601/39.  
 DR N-PSDB; AAT42594.  
 XX  
 XX  
 PT New ligand for receptor type tyrosine kinase - has mol.wt. 22-25  
 PT kilodalton(s) and is positive for Coomassie and PAS staining  
 PS Example 11; Page 44-46; 51pp; Japanese.  
 XX  
 XX This sequence represents the full length receptor-type tyrosine kinase  
 CC protein binding ligand of the invention (also see AA06333 and  
 CC AA06334). The mature protein contained within this sequence contains the  
 CC N-terminal fragment represented by AA06332. The proteins of the  
 CC invention have a molecular weight of 23500 (plus or minus 1500) Da, and  
 CC are positive for Coomassie staining and PAS staining. The proteins of  
 CC the invention bind to the fragment of this protein sequence represented  
 CC by AA06330. The proteins of the invention are new ligands of  
 CC receptor-type tyrosine kinases, and can be prepared by standard  
 CC recombinant techniques.  
 CC  
 XX Sequence 333 AA;  
 SQ  
 Query Match 25.3%; Score 620.5; DB 17; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 4.5e-44;  
 Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;  
 QY 14 GALLILGVLGSLGSLSEPVYVNSANKRFOAGGVLYVPOIGDRLDLCPRARPPGHSS 73  
 DB 14 GVLMLVCRITAKSIVLEPIYVNSNPKFLPGQGLVLPQIGDKLDIICPKV--DSKTV 70  
 QY 74 PNYEFKYLIVGAQGRRCCEAPPAENLLTCDRDLDFETIKFOEYSPNLMGHEFRSH 133  
 DB 71 GQYEVYKVMVDKQADRCITIKENTPLNCAKPPQDIKFTIKFOEYSPNLMGLEFQNK 130  
 QY 134 DYIITATSDGTREGLESLOGGVCLTRGKVTLLRVQ--SPRGGAIVRKVSEMPMER-DR 190  
 DB 131 DYVISTNGSLGELNMGEGVCCQTKMKILMKVGOQDASASASTRKNKPTRRPELAGTN 190  
 QY 191 GAHSLPEKENVLPDPTSNATSGAEGLPPSPMPAVAGAALALLLVAGAGAGNC 250  
 DB 191 GRSSITSPVVKRNPSSSTDSGAGSANNILGSEVALPAGIASGCIITVITLVLL 250  
 QY 251 WRERRAKPSESHPGSGFGRGSLGLGGGGMGRRAEPGELGIALRGGAADPPCFH 310  
 DB 251 KYRRRRHKSPQHTTLLSLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVCFH 303  
 QY 311 YEKVSGDYGHPYIVODGPPQSPNNIY 338  
 DB 304 YEKVSGDYGHPYIVQEMPQSPANNIY 331  
 RESULT 17  
 AA94655  
 ID AA94655 standard; Protein, 333 AA.  
 XX  
 AC AA94655;  
 XX  
 DT 18-OCT-1996 (first entry)  
 XX

DE Ligand for receptor type tyrosine kinase (TK).  
 XX  
 XX Receptor type tyrosine kinase; TK; ligand; differentiation;  
 KW haematopoietic stem cell; tyrosine; bone marrow; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO611212-A1.  
 PN 18-APR-1996.  
 PD  
 XX  
 XX 09-OCT-1995; 95MO-JP02069.  
 PF  
 XX  
 XX 22-DEC-1994; 94JP-0320712.  
 PR 07-OCT-1994; 94JP-024433.  
 PR 26-OCT-1994; 94JP-0262882.  
 XX  
 XX (ASAH ) ASAH KASEI KOGYO KK.  
 PA  
 XX  
 XX Ohno M, Sakano S;  
 PI  
 XX WPI, 1996-209809/21.  
 DR N-PSDB; AAT18395.  
 XX  
 XX  
 XX ligand peptide binding to receptor-type tyrosine kinase - enhances  
 PT intracellular tyrosine phosphorylation, useful for investigation of  
 PT undifferentiated blood cell behaviour  
 XX  
 XX Claim 6; Page 163-164; 193pp; Japanese.  
 XX  
 XX A ligand polypeptide which binds to the extracellular part of a  
 CC specific receptor-type tyrosine kinase and induces phosphorylation  
 CC of tyrosine within the cell can be used in the study of the  
 CC differentiation of blood cells such as the haematopoietic stem  
 CC cells of disease processes such as leukaemia, and of the biology of  
 CC bone marrow transplantation. The ligand plays a role in the  
 CC differentiation process and the specific ligand target is expressed  
 CC in undifferentiated blood cells.  
 CC  
 XX  
 SQ Sequence 333 AA;  
 Query Match 25.3%; Score 620.5; DB 17; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 4.5e-44;  
 Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;  
 QY 14 GALLILGVLGVSGLSLEPYVMNSANKRFQABEGVLYPQIGDRLDLCPRAPPPHSS 73  
 DB 14 GVLWVLCRTAISKSTVLEPIYMNSSNKFPLPGGLVLYPQIGDRLDLCPRAPPPHSS 70  
 QY 74 PNYEFYKLYLVGAQGRRCAPAPNLLTCDRPPDLRFITKFOEYSPNLMGHEFRSH 133  
 DB 71 GQVEYKLYVMVDKQADRCITIKENTPLNCAKPDODIKFTIKFOEYSPNLMGHEFRSH 130  
 QY 134 DYIITATSDGTRREGLESQGVCLTRGMKYLRYGQ--SPRGAVPRKPVSEMPMR-DR 190  
 DB 131 DYIITATSDGTRREGLESQGVCLTRGMKYLRYGQ--SPRGAVPRKPVSEMPMR-DR 190  
 QY 191 GAASLPEKENTLPDPTSNATSRGAEGLPPSPMPAVAGAAGLALLGVAGAGAMC 250  
 DB 191 GRSSTTPPKYKPNKSGSTTDGNSAGHSNMNLGSEVALFAIAGSCITFIYITIVVLL 250  
 QY 251 WRRRRAPSESRHPGSPGFRGSGSLGGCGGMPREAEPEGELIARGGGAADPPCPH 310  
 DB 251 KYRRRHAKHSPQHTTTLSTLAPRKSGNN---NGSEPSDIIIPUR--TADSVFCBH 303  
 QY 311 YEKVSGDYGPVYIVODGPPGSPNIIY 338  
 DB 304 YEKVSGDYGPVYIVQEMPPGSPNIIY 331  
 RESULT 18  
 AAR92743  
 ID AAR92743 standard; Protein; 333 AA.

XX  
 AC AAR92743;  
 XX  
 XX 21-MAY-1996 (first entry)  
 DT  
 XX  
 XX Human hepatoma transmembrane kinase receptor ligand.  
 DE  
 XX Hepatoma transmembrane kinase; Htk; receptor; ligand;  
 KW tyrosine kinase; neurodegenerative disease.  
 KW  
 XX Homo sapiens.  
 OS  
 XX MO9602645-A2.  
 PN 01-FEB-1996.  
 PD  
 XX  
 XX 14-JUL-1995; 95MO-US08812.  
 PF  
 XX  
 XX 20-JUL-1994; 94US-0277722.  
 PR  
 XX (GETH ) GENENTECH INC.  
 PA  
 XX Bennett BD, Matthews W;  
 PI  
 XX WPI, 1996-105907/11.  
 DR N-PSDB; AAT16471.  
 XX  
 XX  
 XX ligand for the hepatoma trans-membrane kinase receptor - useful for  
 PT stimulating and inhibiting cells carrying the receptor, e.g. for  
 PT treating neuro-degenerative disease  
 XX  
 XX Claim 6; Fig 2(A-B); 88pp; English.  
 PS  
 XX Mouse (AAT16470) and human (AAT16471) Htk ligand which bind to, and  
 CC activate, the Htk receptor, have been identified in a variety of  
 CC tissues using a soluble Htk-FC fusion protein.  
 CC The predicted mol.wt. of the murine Htk ligand protein following a  
 CC signal peptide cleavage is 34 kd with an estimated pI of 8.9.  
 CC The murine and human ligands show 96% homology at the amino acid  
 CC level.  
 CC The DNA is used to produce recombinant ligands; for tissue-  
 CC specific typing (partic. as a marker for breast cancer) and as a  
 CC marker for human chromosome 13. The ligands (partic. in soluble  
 CC form) are used to activate the tyrosine kinase domain of the  
 CC Htk receptor, i.e. to stimulate or inhibit growth, differentiation,  
 CC and/or activation of cells contg. the receptor, e.g. treatment  
 CC of neurodegenerative diseases, since they are strongly expressed  
 CC in the cerebral cortex, hippocampus, striatum and cerebellum.  
 CC The ligands are also useful as a control or standard in assays,  
 CC for generation of antibodies, as a mol. wt. marker, for growth  
 CC in vitro of Htk-receptor positive cells, as research agent,  
 CC in screening, etc.  
 CC  
 XX  
 SQ Sequence 333 AA;  
 Query Match 25.3%; Score 620.5; DB 17; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 4.5e-44;  
 Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;  
 QY 14 GALLILGVLGVSGLSLEPYVMNSANKRFQABEGVLYPQIGDRLDLCPRAPPPHSS 73  
 DB 14 GVLWVLCRTAISKSTVLEPIYMNSSNKFPLPGGLVLYPQIGDRLDLCPRAPPPHSS 70  
 QY 74 PNYEFYKLYLVGAQGRRCAPAPNLLTCDRPPDLRFITKFOEYSPNLMGHEFRSH 133  
 DB 71 GQVEYKLYVMVDKQADRCITIKENTPLNCAKPDODIKFTIKFOEYSPNLMGHEFRSH 130  
 QY 134 DYIITATSDGTRREGLESQGVCLTRGMKYLRYGQ--SPRGAVPRKPVSEMPMR-DR 190  
 DB 131 DYIITATSDGTRREGLESQGVCLTRGMKYLRYGQ--SPRGAVPRKPVSEMPMR-DR 190  
 QY 191 GAASLPEKENTLPDPTSNATSRGAEGLPPSPMPAVAGAAGLALLGVAGAGAMC 250  
 DB 191 GRSSTTPPKYKPNKSGSTTDGNSAGHSNMNLGSEVALFAIAGSCITFIYITIVVLL 250



KM insulin; interleukin; fibroblast growth factor; hepatocyte growth factor;  
 KM nerve growth factor; interferon-gamma; tumour necrosis factor; inducer.  
 OS Homo sapiens.  
 XX  
 XX  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /label= signal peptide  
 FT Protein 26..333  
 FT /label= receptor-type tyrosine kinase  
 XX  
 XX JP08280385-A.  
 PN  
 XX 29-OCT-1996.  
 PD  
 XX 13-APR-1995; 95JP-0087878.  
 PF  
 XX 13-APR-1995; 95JP-0087878.  
 PR  
 XX 13-APR-1995; 95JP-0087878.  
 XX  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 XX  
 DR WPI; 1997-014848/02.  
 DR N-PSDB; AAT51236.  
 XX  
 XX LPM cell culture medium - for culturing cells which produce  
 PT receptor-type tyrosine kinase ligand  
 XX  
 PS Disclosure; Page 40-41; 45pp; Japanese.  
 XX  
 XX AAM1307 and AAM1308 represent receptor-type tyrosine kinases. These  
 CC sequences are included in the LPM medium of the invention. The medium  
 CC is used to culture cells capable of producing a ligand for receptor-type  
 CC tyrosine kinase. The ligand serves as a physiologically active  
 CC substance to induce enzyme activity and autophosphorylation of tyrosine  
 CC kinase, which is involved in the development and differentiation of  
 CC animal and insect cells. The ligand can also be used as in a  
 CC pharmaceutical preparation for promoting growth of undifferentiated  
 CC cells. The medium may also include an inducer, such as insulin, an  
 CC interleukin, a fibroblast growth factor, or interferon-gamma.  
 XX  
 SQ Sequence 333 AA;  
 Query Match 25.3%; Score 620.5; DB 18; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 4.5e-44;  
 Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;  
 QY 14 GALLLEVLGLVSGSLSEPYVNSANKRFOAEGGYLYPQIGRLDLCRRAPPGPHSS 73  
 DB 14 GVLMLCRTAISKISVLEPYVNSNSKFLPGQGLVLYPQIGDKLDICPKV---DSKTV 70  
 QY 74 PNEYFYLYVVGAGAGRCAPAPAPNLLTCRPPDLRTIKFOEYSPYLMGHEFRSH 133  
 DB 74 GQYEVYVYVVDQDQDRCTIKENTPLNCAFPDODIKTIKFOEYSPYLMGHEFRSH 130  
 QY 134 DYIITATSDGTREGLESLOGGVCITRGMKYLVRVQ--SPRGAVPRKPYSEMPMER-DR 190  
 DB 131 DYIITATSDGTREGLESLOGGVCITRGMKYLVRVQ--SPRGAVPRKPYSEMPMER-DR 190  
 QY 131 DYIITATSDGTREGLESLOGGVCITRGMKYLVRVQ--SPRGAVPRKPYSEMPMER-DR 190  
 DB 131 DYIITATSDGTREGLESLOGGVCITRGMKYLVRVQ--SPRGAVPRKPYSEMPMER-DR 190  
 QY 191 GAHSLPEKGENLPDGTNATSRGAGRPSPSMRAVNAAGGALLLLGVAGAGAAC 250  
 DB 191 GRSTTSFYKPNPENGSTIDGNSAGHSNNITLSEVALFAIAGCITFTVITITLVLL 250  
 QY 251 WRRRRAPSSSRHPGSPGSGSLGIGGGGMPREAEFGELIALRGGAADPPFCFH 310  
 DB 251 KYRRRRHKHSPOHTTTLSLSTLATPKRSGNN---NGSEPSDITITPLR---TDSVFCFH 303  
 QY 311 YEKYSGLYGHPIVYVODGPPQSPNITY 338  
 DB 304 YEKVSGDYGHPIVYVODGPPQSPNITY 331

RESULT 21  
 AAM06334

ID AAM06334 standard; protein; 308 AA.  
 XX  
 AC AAM06334;  
 XX  
 XX  
 DT 17-JAN-1997 (first entry)  
 XX  
 XX  
 DE Ligand #2 for receptor-type tyrosine kinase protein.  
 XX  
 KM Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;  
 KM human.  
 XX  
 XX  
 OS Homo sapiens.  
 XX  
 XX JP08188596-A.  
 PN  
 XX 23-JUL-1996.  
 PD  
 XX 13-JAN-1995; 95JP-0003677.  
 PF  
 XX 09-NOV-1994; 94JP-0275411.  
 PR 19-OCT-1994; 94JP-0253848.  
 XX  
 PA (ASAH ) ASAH KASEI KOGYO KK.  
 XX  
 DR WPI; 1996-388601/39.  
 XX  
 XX New ligand for receptor type tyrosine kinase - has mol.wt. 22-25  
 PT kilo: dalton(s) and is positive for Coomassie and PAS staining  
 XX  
 PS Claim 7; Page 44; 51pp; Japanese.  
 XX  
 XX AAM06333 and AAM06334 represent receptor-type tyrosine kinase receptor  
 CC binding ligands of the invention. These sequences both have the  
 CC N-terminal sequence represented by AAM06332. The ligands of the  
 CC invention recognize the fragment of the receptor type kinase receptor  
 CC represented by AAM06330 (see AAM06331 for full length sequence). The  
 CC proteins of the invention have a molecular weight of 23500 (plus or minus  
 CC 1500) Da, and are positive for Coomassie staining and PAS staining. The  
 CC protein is a new ligand of receptor-type tyrosine kinases, and can be  
 CC prepared by standard recombinant techniques.  
 XX  
 SQ Sequence 308 AA;  
 Query Match 24.9%; Score 610.5; DB 17; Length 308;  
 Best Local Similarity 41.7%; Pred. No. 2.8e-43;  
 Matches 130; Conservative 48; Mismatches 121; Indels 13; Gaps 5;  
 QY 30 LEPYVNSANKRFOAEGGYLYPQIGRLDLCRRAPPGPHSSPNYFYLYVVGAGG 89  
 DB 5 LEPIYVNSNSKFLPGQGLVLYPQIGDKLDICPKV---DSKTVGQYEVYVYVVDQD 61  
 QY 90 RCCEAPAPAPNLLTCRPPDLRTIKFOEYSPYLMGHEFRSHDYIITATSDGTREG 149  
 DB 62 DRCITIKENTPLNCAFPDODIKTIKFOEYSPYLMGHEFRSHDYIITATSDGTREG 121  
 QY 150 SLOGGVCITRGMKYLVRVQ--SPRGAVPRKPYSEMPMER-DRGAHSLPEKGENLPD 206  
 DB 122 NOEGGVCOTTRAMKILMKYGDASAGSTRNKDPTRRRELZAGTNGRSTTSFYKAPPGS 181  
 QY 207 PTSNATSRGAGRPSPSMRAVNAAGGALLLLGVAGAGAACWRRRRAPSSSRHPG 266  
 DB 182 STDGNSAGHSNNITLSEVALFAIAGCITFTVITITLVLLKTYRRRRHKHSPOHTT 241  
 QY 267 GSFGRGSLGIGGGGMPREAEFGELIALRGGAADPPFCFHYEKVSGDYGHPIVY 326  
 DB 242 LSLSTLATPKRSGNN---NGSEPSDITITPLR---TDSVFCFHYEKVSGDYGHPIVY 294  
 QY 327 DGPPQSPNITY 338  
 DB 295 EMPQSPNITY 306

RESULT 22



CC Elk-12, EFL-6, ELF-3 and LERK-8.  
 XX Sequence 89 AA;  
 SQ  
 Query Match 19.7%; Score 483; DB 21; Length 89;  
 Best Local Similarity 97.7%; Pred. No. 3,4e-33;  
 Matches 85; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 252 RRRRAKSESHHPGSGFGRGSLGLGGGGMGPRAEPGELGIALRGGAADPPFCFHY 311  
 DB 1 RRRRAKSESHHPGSGFGRGSLGLGGGGMGPRAEPGELGIALRGGAADPPFCFHY 60  
 QY 312 EKVSQDYGHPIYIVODGPPGSPNNIYY 338  
 DB 61 EKVSQDYGHPIYIVODGPPGSPNNIYY 87  
 RESULT 24  
 ID AAY6782 standard; Protein: 658 AA.  
 XX  
 AC AAY6782;  
 XX  
 DT 26-SEP-2000 (first entry)  
 XX  
 DE Ephrin-B2-Ephrin-B2-FC fusion protein.  
 XX  
 KW Ephrin-B2; ELK receptor; ligand; dimer; Fc domain; fusion protein;  
 KM Efl-6 antagonist; neurological.  
 XX  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Synthetic.  
 XX  
 FH Key  
 FT Peptide 1..30 Location/Qualifiers  
 FT /label= Signal peptide  
 FT /note= "Derived from ephrin-B2 ectodomain 1"  
 FT Protein 31..225  
 FT /label= Ephrin-B2\_Ectodomain-1  
 FT Peptide 226..228  
 FT /label= Bridging\_peptide  
 FT Protein 229..423  
 FT /label= Ephrin-B2\_ectodomain\_2  
 FT Peptide 424..426  
 FT /label= Bridging\_peptide  
 FT Region 427..658  
 FT /label= Human\_IgG1\_Fc\_region  
 FT  
 XX  
 PN W0200037642-A1.  
 XX  
 PD 29-JUN-2000.  
 XX  
 PF 23-DEC-1999; 99WO-US30900.  
 XX  
 PR 23-DEC-1998; 98US-0113387.  
 XX  
 PA (REGG-) REGENERON PHARM INC.  
 XX  
 PI Davis SJ, Gale NM, Yancopoulos GD, Stahl N;  
 XX  
 DR WPI; 2000-442670/38.  
 DR N-PSDB; AAA51346.  
 XX  
 PT Polynucleotide encoding a fusion polypeptide, useful for promoting  
 PT differential function and influencing phenotype, comprises two subunits  
 PT containing at least one copy of the receptor binding domain of a ligand  
 XX  
 PS Example 12; Fig 15A-E; 97pp; English.  
 XX  
 CC Production of homogenous forms of clustered ligands is broadly applicable  
 CC to improve the affinity and/or increase the activity of a ligand as  
 CC compared to the native form of the ligand. Ephrin fusion proteins have  
 CC been constructed, which may be useful for treating neurological

CC disorders. The ephrin fusion proteins are preferably capable of binding  
 CC to Elk receptor and are especially Efl-6 antagonists. The fusion proteins  
 CC were constructed after it was demonstrated that similar improved  
 CC activities could be achieved using Tie-2 receptor ligands.  
 CC Angiopoietin-1 (Ang-1) is a Tie-2 receptor ligand and is an agonist for  
 CC Tie-2, whereas angiopoietin-2 (Ang-2) is a naturally occurring antagonist  
 CC of the Tie-2 receptor. The fibronogen domains (FD) of Ang-1 and Ang-2 are  
 CC the receptor-binding domains and dimerized versions, of e.g. Ang-1-FD-Fc  
 CC (Ang-1 fibronogen domain fused to an Fc domain), can bind to the Tie-2  
 CC receptor with much higher affinity than monomeric Ang-1-FD (dimerization  
 CC occurs between the Fc components of adjacent molecules). However,  
 CC Ang-1-FD-Fc is not able to induce phosphorylation (activate) the Tie-2  
 CC receptor on endothelial cells unless it is further clustered with goat  
 CC anti-human Fc antibodies. The novel fusion proteins, mutant versions of  
 CC Ang-1-FD and Ang-2-FD, were designed that were intrinsically more highly  
 CC clustered. Tie-2 agonist fusion proteins may be used as haematopoietic  
 CC factors. Tie-2 receptor antagonist fusion proteins may be used to  
 CC diagnose or treat, e.g. myeloproliferative or other proliferative  
 CC disorders of blood forming organs, e.g. thrombocythemiae, polycythemiae  
 CC and leukemias.  
 CC  
 XX  
 SQ Sequence 658 AA;  
 XX  
 Query Match 18.7%; Score 458.5; DB 21; Length 658;  
 Best Local Similarity 30.3%; Pred. No. 5e-30;  
 Matches 128; Conservative 40; Mismatches 111; Indels 143; Gaps 12;  
 QY 5 HSGPGGVRVALLLLIGLVGSLSEBPVYNSANKRFQAGGIVLYPOIDRDLDCR 64  
 DB 218 HSGNN-----LLGCGPIV---LEPIYNNSSNSKFLPGQGLVLYPOIDRDLDCR 265  
 QY 65 ARPPGHSSPNVEFYKLVYVGAQGRCEADPPAPVLLTCDRPPDLDFETIKFOEYSNL 124  
 DB 266 V---DSKTVGQYEVYKVMWDKQADRCTIKENTPLNCAKRPDQDVFTIKFGERSNL 322  
 QY 125 WGHFRRSHDYYIIATSDTREGLESLOGVCLTRGMVLLRVGQSPRGAVPRKPVSEM 184  
 DB 323 WGLEFGKNDYYIIISTNSGLELDNQGVCQTRAMKILMKVGQD----- 368  
 QY 185 PMERDRGAHSLSPKENVLPDPTSNATSRGAEGLPPSPMAYAGAGLALLLVGAV 244  
 DB 369 -----ASSAGSANNHGTREPLE-----AG 389  
 QY 245 AGGAMCWRRRAKPSSESHHPGSGFGRGSLGLGGGGMGPRAEPGELGIALRGGAAD 304  
 DB 390 TNG-----RSSTSPFVKENPGSSTDGNSAGHSGNNLL-----GGGGE 427  
 QY 305 P-----PCPHYEVKYSG-----DYGH----- 320  
 DB 428 PKSCDKTHTCPCPAPBELLGSPVFLPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFN 487  
 QY 321 -----PYIVODGPPGSPSPNYYTTSISVLEWPIILHTIQLFFMRKSGRVTTFLPPQVI 374  
 DB 488 WYVDGVENVNAKTRREBYNSTYRVSVL--TVLHQWLNQKKEKVC-KVSKKLLPAPIE 544  
 QY 375 TT 376  
 DB 545 KT 546  
 RESULT 25  
 ID AAY6781 standard; Protein: 683 AA.  
 XX  
 AC AAY6781;  
 XX  
 DT 26-SEP-2000 (first entry)  
 XX  
 DE Ephrin-B1-Ephrin-B1-FC fusion protein.  
 XX  
 KW Ephrin-B1; ELK receptor; ligand; dimer; Fc domain; fusion protein;  
 KM Efl-6 antagonist; neurological.  
 XX



OS Chimeric - Homo sapiens.  
 XX Chimeric - Synthetic.  
 FT Key Location/Qualifiers  
 FT Peptide 1..29  
 FT /label= Signal\_peptide  
 FT /note= "derived from ELK-L Ectodomain 1"  
 FT Protein 30..237  
 FT /label= ELK-L\_ectodomain\_1  
 FT Peptide 238..240  
 FT /label= Bridging\_peptide  
 FT Protein 241..448  
 FT /label= ELK-L\_Ectodomain\_2  
 FT Peptide 449..451  
 FT /label= Bridging\_peptide  
 FT Region 452..683  
 FT /label= Human\_IgG1\_Fc\_region  
 XX WO200037642-A1.  
 XX 29-JUN-2000.  
 XX 23-DEC-1999; 99WO-US30900.  
 XX 23-DEC-1998; 98US-0113387.  
 XX (REGGE-) REGENERON PHARM INC.  
 XX Davis SJ, Gale NW, Yancopoulos GD, Stahl N;  
 XX WPI; 2000-442670/38.  
 XX N-PSDB; AAA51345.  
 PT Polynucleotide encoding a fusion polypeptide, useful for promoting  
 PT differential function and influencing phenotype, comprises two subunits  
 PT containing at least one copy of the receptor binding domain of a ligand  
 XX  
 PS Example 12; Fig 14A-E; 97pp; English.  
 CC Production of homogeneous forms of clustered ligands is broadly applicable  
 CC to improve the affinity and/or increase the activity of a ligand as  
 CC compared to the native form of the ligand. Ephrin fusion proteins have  
 CC been constructed, which may be useful for treating neurological  
 CC disorders. The ephrin fusion proteins are preferably capable of binding  
 CC to Etk receptor and are especially Etk-6 antagonists. The fusion proteins  
 CC were constructed after it was demonstrated that similar improved  
 CC activities could be achieved using Tie-2 receptor ligands.  
 CC Angiopoietin-1 (Ang-1) is a Tie-2 receptor ligand and is an agonist for  
 CC Tie-2, whereas angiopoietin-2 (Ang-2) is a naturally occurring antagonist  
 CC of the Tie-2 receptor. The fibrinogen domains (FD) of Ang-1 and Ang-2 are  
 CC the receptor-binding domains and dimerized versions, of e.g. Ang-1-FD-Fc  
 CC (Ang-1 fibrinogen domain fused to an Fc domain), can bind to the Tie-2  
 CC receptor with much higher affinity than monomeric Ang-1-FD (dimerization  
 CC occurs between the Fc components of adjacent molecules). However,  
 CC Ang-1-FD-Fc is not able to induce phosphorylation (activate) the Tie-2  
 CC receptor on endothelial cells unless it is further clustered with goat  
 CC anti-human Fc antibodies. The novel fusion proteins, mutant versions of  
 CC Ang-1-FD and Ang-2-FD, were designed that were intrinsically more highly  
 CC clustered. Tie-2 agonist fusion proteins may be used as haematopoietic  
 CC factors. Tie-2 receptor antagonist fusion proteins may be used to  
 CC diagnose or treat, e.g. myeloproliferative or other proliferative  
 CC disorders of blood forming organs, e.g. thrombocythemia, polycythemia  
 CC and leukemias.  
 CC  
 XX Sequence 683 AA;  
 SQ  
 Query Match 18.5%; Score 454; DB 21; Length 683;  
 Best Local Similarity 44.0%; Pred. No. 1.3e-29;  
 Matches 92; Conservative 32; Mismatches 72; Indels 13; Gaps 3;

QY 87 AGGRCAPAPNLTLCDDPDLFTTKFOEYSPNLWGHEFRSHDXYIATSDGRE 146  
 DB 295 EQAAACSTVLDNRVAVTCRPFQEIHTTKFOEYSPNKGLEFKGHDIYISTNGSLE 354  
 QY 147 GLESLGGVCLTRGKMKVLLRVGQSPRGAVPRKPYSEMEREDRGAASLE-PGKENLPG 205  
 DB 355 GLENREGVCFRTTKMKIKVGDPAVAVPEQLTTSRPSKADNVTVMATQAPGSRGSLG 414  
 QY 206 DPTSNATSRGAGPLPPEPMVAAGAGG 234  
 DB 415 DSDGKHEITVNGEKGSP-----GASGG 436

RESULT 26  
 AAE24019  
 ID AAE24019 standard; Protein; 229 AA.

AC AAE24019;  
 DT 23-SEP-2002 (first entry)  
 XX Murine ephrin B2 ligand.

XX Murine; extracellular region; Eph B receptor; ephrin B ligand; tumour;  
 XX ocular neovascularisation; gene therapy; cytostatic.

OS Mus musculus.

PN WO200226827-A1.

PD 04-APR-2002.

PF 28-SEP-2001; 2001WO-EP11252.

XX 29-SEP-2000; 2000CH-0001910.

XX (NOVS ) NOVARTIS AG.

PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Martiny-Baron G, Wood JM, Liao G;

DR WPI; 2002-405047/43.

PT N-PSDB; AAD38768.

XX New soluble polypeptides of the extracellular region of the Eph B  
 PT receptor and ephrin B ligands, useful in gene therapy for treating a  
 PT tumor disease or ocular neovascularization in mammals, including humans

XX Claim 6; Page 48-49; 55pp; English.

XX The invention relates to new isolated soluble polypeptides comprising an  
 CC amino acid sequence of an extracellular region of an Eph B receptor or an  
 CC amino acid sequence of an ephrin B ligand, or essentially similar amino  
 CC acid sequences of extracellular region of an Eph B receptor or an amino  
 CC acid sequence of an ephrin B ligand. Extracellular region of an Eph B  
 CC receptor is capable of binding with high affinity to a mammalian ephrin B  
 CC ligand. Ephrin B ligand is capable of binding with high affinity to a  
 CC mammalian Eph B receptor. The nucleic acid encoding the polypeptides or  
 CC the polypeptides are useful for producing pharmaceutical compositions for  
 CC treating a tumour disease or ocular neovascularisation in mammals,  
 CC including humans. The vector is useful in producing a pharmaceutical  
 CC composition for treating the diseases cited above in mammals,  
 CC particularly humans, by means of gene therapy. The present sequence is  
 CC murine ephrin B2 ligand.

XX Sequence 229 AA;

SQ

Query Match 18.2%; Score 447; DB 23; Length 229;  
 Best Local Similarity 36.4%; Pred. No. 1.2e-29;  
 Matches 100; Conservative 33; Mismatches 80; Indels 62; Gaps 4;

QY 14 GALLILGLVGLVSGLSLEPYVWNSANKRFQAEGLVLYPQIGRLDLICPPARPPGPHSS 73  
 Db 17 GLIMVLCRTAISRSLVLEPIYWNSSNGKFLPGQGLVLYPQIGRLDLICPKV---DSKTV 73  
 QY 74 PNEYFYKLYLVGAAGRCCEAPAPNNLLTCRDPDLRFITKFOEYSPNLMGHEFRSH 133  
 Db 74 GQYFYKYVMDKQADRCITIKENTPLNCAKPDQIVKTIKFQETSPNLMGLEFQK 133  
 QY 134 DYYIATSDGTREGLESIQGVCLTRGMKVLRLVQSPRGAVPRKPVSEMERDRGAA 193  
 Db 134 DYYIISTNSGSLGLEIDNOEGGVCTQTRAMKILMKYGO----- 170  
 QY 194 HSLERPKENLPQDPTSNATSRGAEGLPPSPMPAVAGAAGLALLLVAGAGAMCMRR 253  
 Db 171 -----ASSAGSRNKHGPTRRPELE-----AGTNG----- 194  
 QY 254 RRAKPSERHPGPGSFRGSGSLGLGGGGMGPREA 288  
 Db 195 RSTSTSPFVKENPGSSTDGNSAGHSGNNLLGSEVA 229  
 RESULT 27  
 AAE24020  
 ID AAE24020 standard; Protein; 229 AA.  
 AC AAE24020;  
 XX 23-SEP-2002 (first entry)  
 DT Human ephrin B2 ligand.  
 DE Human ephrin B2 ligand.  
 KM Human; extracellular region; Eph B receptor; ephrin B ligand; tumour;  
 KW ocular neovascularisation; gene therapy; cytostatic.  
 OS Homo sapiens.  
 XX WO200226827-A1.  
 PN 04-APR-2002.  
 PD 28-SEP-2001; 2001WO-EPI1252.  
 PF 29-SEP-2000; 2000CH-0001910.  
 PR 29-SEP-2000; 2000CH-0001910.  
 XX (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.  
 PI Martiny-baron G, Wood JM, Lieu G,  
 XX WPI; 2002-405047/43.  
 DR N-PSDB; AAD38769.  
 XX  
 PT New soluble polypeptides of the extracellular region of the Eph B  
 PT receptor and ephrin B ligands, useful in gene therapy for treating a  
 PT tumor disease or ocular neovascularization in mammals, including humans  
 PS  
 PS Claim 6; Page 51-52; 55pp; English.  
 CC The invention relates to new isolated soluble polypeptides comprising an  
 CC amino acid sequence of an ephrin B ligand, or essentially similar amino  
 CC amino acid sequence of an ephrin B ligand, or essentially similar amino  
 CC acid sequence of an ephrin B ligand. Extracellular region of an Eph B  
 CC receptor is capable of binding with high affinity to a mammalian ephrin B  
 CC ligand. Ephrin B ligand is capable of binding with high affinity to a  
 CC mammalian Eph B receptor. The nucleic acid encoding the polypeptides or  
 CC the polypeptides are useful for producing pharmaceutical compositions for  
 CC treating a tumour disease or ocular neovascularisation in mammals,  
 CC including humans. The vector is useful in producing a pharmaceutical  
 CC composition for treating the diseases cited above in mammals,  
 CC particularly humans, by means of gene therapy. The present sequence is  
 CC human ephrin B2 ligand.

XX SQ Sequence 229 AA;  
 Query Match 18.1%; Score 443; DB 23; Length 229;  
 Best Local Similarity 35.6%; Pred. No. 2,7e-29;  
 Matches 98; Conservative 36; Mismatches 79; Indels 62; Gaps 4;  
 QY 14 GALLILGLVGLVSGLSLEPYVWNSANKRFQAEGLVLYPQIGRLDLICPPARPPGPHSS 73  
 Db 14 GLIMVLCRTAISRSLVLEPIYWNSSNGKFLPGQGLVLYPQIGRLDLICPKV---DSKTV 70  
 QY 74 PNEYFYKLYLVGAAGRCCEAPAPNNLLTCRDPDLRFITKFOEYSPNLMGHEFRSH 133  
 Db 71 GQYFYKYVMDKQADRCITIKENTPLNCAKPDQIVKTIKFQETSPNLMGLEFQK 130  
 QY 134 DYYIATSDGTREGLESIQGVCLTRGMKVLRLVQSPRGAVPRKPVSEMERDRGAA 193  
 Db 131 DYYIISTNSGSLGLEIDNOEGGVCTQTRAMKILMKYGO----- 166  
 QY 194 HSLERPKENLPQDPTSNATSRGAEGLPPSPMPAVAGAAGLALLLVAGAGAMCMRR 253  
 Db 167 -----DASAGSTRNKDPTRRPELE-----AGTNG----- 191  
 QY 254 RRAKPSERHPGPGSFRGSGSLGLGGGGMGPREA 288  
 Db 192 RSTSTSPFVKENPGSSTDGNSAGHSGNNLLGSEVA 226  
 RESULT 28  
 AAM06333  
 ID AAM06333 standard; protein; 195 AA.  
 AC AAM06333;  
 XX 17-JAN-1997 (first entry)  
 DT Ligand #1 for receptor-type tyrosine kinase protein.  
 DE Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;  
 KW human.  
 OS Homo sapiens.  
 XX JP08186596-A.  
 PN 23-JUL-1996.  
 PD 13-JAN-1995; 95JP-0003677.  
 PF 09-NOV-1994; 94JP-0275411.  
 PR 19-OCT-1994; 94JP-0253848.  
 XX (ASAH ) ASAH KASEI KOGYO KK.  
 PA WPI; 1996-388601/39.  
 DR  
 XX  
 PT New ligand for receptor type tyrosine kinase - has mol.wt. 22-25  
 PT kilo: dalton(s) and is positive for Coomassie and PAS staining  
 PS  
 PS Claim 5; Page 43; 51pp; Japanese.  
 CC AAM06333 and AAM06334 represent receptor-type tyrosine kinase receptor  
 CC binding ligands of the invention. These sequences both have the  
 CC N-terminal sequence represented by AAM06332. The ligands of the  
 CC invention recognise the fragment of the receptor type kinase receptor  
 CC represented by AAM06330 (see AAM06331 for full length sequence). The  
 CC proteins of the invention have a molecular weight of 23500 (plus or minus  
 CC 1500) Da, and are positive for Coomassie staining and PAS staining. The  
 CC protein is a new ligand of receptor-type tyrosine kinases, and can be  
 CC prepared by standard recombinant techniques.  
 XX Sequence 195 AA;



[illegible]

Query Match	Best Local Similarity	Score	DB	Length
Matches 52; Conservative 15; Mismatches 24; Indels 5; Gaps 1;	11.64;	284.5;	23;	92;
42 FOAEGGYLVYEQIGRDLICPRARPPGHSPNYEFKYLIVGAGGRCRAPPNLL 101	1	FLSGGGLVYIKIGDKDILICPRALAGRP-----YEVKILIVAPPCQAALSTYLDPRVL 55		
102 LTCRPPDIDLRFITIKFQEPYSPNLWGHEFRSHDYIT 137		VTGNRPGEIRFTTIKFQFSPNYWGLERKCKHDYVI 91		
AAAM37534	AAAM37534 standard; Protein; 136 AA.			
AAAM37534;				
17-OCT-2001 (first entry)				
Peptide #11571 encoded by probe for measuring placental gene expression.				
Probe; microarray; human; placenta; antenatal diagnosis; genetic disorder.				
Homo sapiens.				
MO200157272-A2.				
09-AUG-2001.				
30-JAN-2001; 2001WO-US00663.				
04-FEB-2000; 2000US-0180312.				
26-MAY-2000; 2000US-0207456.				
30-JUN-2000; 2000US-0608408.				
03-AUG-2000; 2000US-0632366.				
21-SEP-2000; 2000US-0234687.				
27-SEP-2000; 2000US-0236359.				
04-OCT-2000; 2000GB-0024263.				
(MOLE-) MOLECULAR DYNAMICS INC.				
Penn SG, Hanzel DK, Chen W, Rank DR,				
WPI, 2001-488897/53.				
Human genome-derived single exon nucleic acid probes useful for				
analyzing gene expression in human placenta -				

PS Claim 27; SEQ ID No 37803; 654bp; English.  
 CC The present invention relates to single exon nucleic acid probes (SENP:  
 CC see Aa13115-Aa157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX  
 SQ Sequence 136 AA;  
 Query Match 8.0%; Score 196.5; DB 22; Length 136;  
 Best Local Similarity 36.8%; Pred. No. 8.8e-09;  
 Matches 50; Conservative 13; Mismatches 34; Indels 39; Gaps 5;  
 QY 230 GAAGGAA-----LLLGAGAGGA-----MCMRRBRAPSESRRHGP 266  
 DB 11 GASGGSSGDDPGRFNSKALFAAVGAGCVIFLLITFLTVLLIKRKRRKHTQO----- 65  
 QY 267 GSFGRGSGTGL---GGGGGMPREAPPELIGALRGGAADPPFCPHYEKVSGDYGHV 322  
 DB 66 ----RAAALSLSTLASPKGSGTAGTEPSDITILPR---TTENNYCPHYEKVSGDYGHV 118  
 QY 323 YIVODGPPSPENITY 338  
 DB 119 YIVQEMPQSPANITY 134  
 RESULT 33  
 ABG46394  
 ID ABG46394 standard; Peptide: 136 AA.  
 AC ABG46394;  
 XX  
 DT 19-AUG-2002 (first entry)  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 36059.  
 XX  
 KM Human; single exon probe; asthma; lung cancer; COPD, ILD,  
 KM chronic obstructive pulmonary disease; interstitial lung disease;  
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KM Heremansky-Pudlak syndrome; sarcoidosis; pulmonary haemangioendothelioma;  
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;  
 KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KM primary ciliary dyskinesia; pulmonary hypertension;  
 KM hyaline membrane disease.  
 KM  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-063236P.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX

PS Claim 27; SEQ ID No 36059; 634bp; English.  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridize at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung; comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA; and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridization of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridization to a single exon  
 CC microarray having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer; chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Heremansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemangioendothelioma, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 136 AA;  
 Query Match 8.0%; Score 196.5; DB 23; Length 136;  
 Best Local Similarity 36.8%; Pred. No. 8.8e-09;  
 Matches 50; Conservative 13; Mismatches 34; Indels 39; Gaps 5;  
 QY 230 GAAGGAA-----LLLGAGAGGA-----MCMRRBRAPSESRRHGP 266  
 DB 11 GASGGSSGDDPGRFNSKALFAAVGAGCVIFLLITFLTVLLIKRKRRKHTQO----- 65  
 QY 267 GSFGRGSGTGL---GGGGGMPREAPPELIGALRGGAADPPFCPHYEKVSGDYGHV 322  
 DB 66 ----RAAALSLSTLASPKGSGTAGTEPSDITILPR---TTENNYCPHYEKVSGDYGHV 118  
 QY 323 YIVODGPPSPENITY 338  
 DB 119 YIVQEMPQSPANITY 134  
 RESULT 34  
 AAY71437  
 ID AAY71437 standard; peptide: 82 AA.  
 AC AAY71437;  
 XX  
 DT 04-OCT-2000 (first entry)  
 DE Human ephrin B2 C-terminal cytoplasmic domain.  
 KM PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
 KM

KM cellular processes; nerve cell interaction; regeneration of nerve cell;  
 KM axonogenesis; antiproliferative; proliferative disorder; treatment;  
 KM differentiative disorder; human; cytoplasmic domain.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Region 50..68 /note= "Conserved region"

FT Modified-site 53 /note= "Conserved tyrosine phosphorylation site"

FT Modified-site 60 /note= "Conserved tyrosine phosphorylation site"

FT Modified-site 65 /note= "Conserved tyrosine phosphorylation site"

FT Region 71..75 /note= "Conserved tyrosine phosphorylation site"

FT Region 77..82 /note= "Conserved region"

FT Domain 79..82 /note= "Conserved region"

FT Modified-site 79 /label= PDZ\_domain

FT Modified-site 80 /note= "Conserved tyrosine phosphorylation site"

FT Modified-site /note= "Conserved tyrosine phosphorylation site"

XX MO200031124-A2.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-CA01101.

PR 20-NOV-1998; 98US-0109158.

PA (MOUN ) MOUNT SINAI HOSPITAL.

PI Lin D, Pawsen A;

DR WPI; 2000-400038/34.

XX Isolated complex for creating proliferative or differentiative disorders comprises B class ephrin and PDZ domain containing protein -

XX Example; Fig 1; 59pp; English.

XX The patent discloses a complex comprising of a B class ephrin and PDZ domain containing protein. B class ephrins function as ligands for Eph receptor tyrosine kinases (RTK) and possess a transmembrane element and a highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ binding site. This complex is used in methods to modulate the interaction of a B class ephrin and PDZ domain containing protein and to identify modulator compounds. It is also used for modulating cellular processes like, axonogenesis, nerve cell interactions and regeneration of nerve cells. The complex is also useful for treating proliferative or differentiative disorders associated with this protein complex.

XX The present sequence is the human ephrin B2 C-terminal cytoplasmic domain, comprising conserved sequences reminiscent of known or predicted binding sites for PDZ domains. Ephrin B2 is also known as Htk-L, Efr-2, LERK-5 and NLERK-1.

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

ID AAB54187 standard; Protein; 106 AA.

XX AAB54187;

AC AAB54187;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:639.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; cytostatic; neuroprotective; neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological; antiinflammatory; cardiant; gene therapy; chromosome mapping; linkage analysis; tissue identification; tissue typing; forensic; neural; immune system; muscular; reproductive; gastrointestinal; pulmonary; cardiovascular; renal; proliferative.

XX Homo sapiens.

OS WO200055320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05989.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-576444/54.

XX N-PSDB; AAC98952.

XX New nucleic acid that is a pancreatic cancer antigen for preventing, treating, or ameliorating a medical condition, particular pancreatic cancer, or for use in assays for diagnosing a pathological condition -

XX Claim 11; Page 1078; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated proteins, called pancreatic cancer antigens, given in AAB54008 to AAB54466. The human pancreatic cancer antigens have cytostatic, neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive, gynaecological, cardiant and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer.

XX Agonists and antagonists to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridization probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent sequences used in the exemplification of the present invention.

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

Query Match 7.9%; Score 193.5; DB 21; Length 106;

Best Local Similarity 42.6%; Pred. No 1.2e-08;

Matches 46; Conservative 11; Mismatches 26; Indels 25; Gaps 4;

QY 235 TALLTGGVAGAGAMCWRRAKPSRHPGSGFGSGTGL---GGGGMGPREAP 290

DB 18 LTVLLKL-----RRRRKHTQ-----RAAALSLSTLASPKGSGSGTAGTEP 59

QY 291 GELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

QY 279 GGGGNGPREAPGELGIALRGGAADPPCPHKEKVSQDYGHPVIYQDPPSPPIIY 338

DB 27 GANNNG---SEPSDVIPILR---TADSVFCPHYKVGSDYGHPIYIVQEMPSPANIY 80

Db 60 SDIIILRLR---TTENNYCPHYEKVSGDYGHPVTVQEMPPQSPANITY 104

RESULT 36  
AAV71436  
ID AAV71436 standard; peptide; 82 AA.  
XX  
AC AAV71436;  
DT 04-OCT-2000 (first entry)  
XX  
DE Human ephrin B1 C-terminal cytoplasmic domain.  
XX  
KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
KW cellular process; nerve cell interaction; regeneration of nerve cell;  
KW axonogenesis; antiproliferative; proliferative disorder; treatment;  
KW differentiative disorder; human; cytoplasmic domain.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 50..68  
FT Modified-site /note= "Conserved region"  
FT Modified-site 53 /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site 60 /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site 65 /note= "Conserved tyrosine phosphorylation site"  
FT Region 71..75 /note= "Conserved tyrosine phosphorylation site"  
FT Region /note= "Conserved region"  
FT Region 77..82 /note= "Conserved region"  
FT Domain 79..82 /label= PDZ\_domain  
FT Modified-site 79 /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site 80 /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site /note= "Conserved tyrosine phosphorylation site"  
XX  
PN MO200031124-A2.  
XX  
PD 02-JUN-2000.  
XX  
PF 19-NOV-1999; 99WO-CA01101.  
XX  
PR 20-NOV-1998; 98US-0109158.  
XX  
PA (MOUNT SINAI HOSPITAL.  
XX  
PI Lin D, Pawson A;  
XX  
DR WPI; 2000-400038/34.  
XX  
PT Isolated complex for treating proliferative or differentiative  
XX disorders comprises B class ephrin and PDZ domain containing protein -  
XX  
PS Example; Fig 1; 59pp; English.  
XX  
CC The patent discloses a complex comprising of a B class ephrin and PDZ  
CC domain containing protein. B class ephrins function as ligands for Eph  
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a  
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ  
CC binding site. This complex is used in methods to modulate the interaction  
CC of a B class ephrin and PDZ domain containing protein and to identify  
CC modulator compounds. It is also used for modulating cellular processes  
CC like, axonogenesis, nerve cell interactions and regeneration of nerve  
CC cells. The complex is also useful for treating proliferative or  
CC differentiative disorders associated with this protein complex.  
CC The present sequence is the human ephrin B1 C-terminal cytoplasmic  
CC domain, comprising conserved sequences reminiscent of known or predicted  
CC binding sites for PDZ domains. Ephrin B1 is also known as IERK-2, EIk-L,

CC EFL-3, Cdk-L and STRAL.  
XX  
SQ Sequence 82 AA;

Query Match 7.8%; Score 192; DB 21; Length 82;  
Best Local Similarity 46.2%; Pred. No. 1,1e-08;  
Matches 42; Conservative 9; Mismatches 24; Indels 16; Gaps 3;

OY 252 RRRAPKPSERHPEPGSGFGGSLGL---GGGGGNGPREAPGELGIALRGGAADPPF 307  
DB 2 RKRHRKHTQQ-----FAALSLSTLASPKGSGTAGTEPSDIIILRLR---TTENNY 49

OY 308 CPHYEKVSGDYGHPVTVQDPPQSPENITY 338  
DB 50 CPHYEKVSGDYGHPVTVQEMPPQSPANITY 80

RESULT 37  
ABBS8621  
ID ABBS8621 standard; Protein; 652 AA.  
XX  
AC ABBS8621;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 2655.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN MO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li FMD, Myers EW;  
XX  
DR WPI; 2001-6556860/75.  
XX  
DR N-PADB; ABL02724.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
PS Disclosure; SEQ ID NO 2655; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABJ30511), expressed DNA  
XX sequences (ABJ1840-ABJ16175) and the encoded proteins  
XX (ABBS7737-ABBS72072).  
XX  
CC The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 652 AA;

Query Match 7.8%; Score 192; DB 22; Length 652;  
Best Local Similarity 21.2%; Pred. No. 1.6e-07;  
Matches 103; Conservative 64; Mismatches 154; Indels 164; Gaps 22;  
OY 33 VYNSANKRQAGG-GYLVYFQIG-----DRIDLCPAPRPPGPHSSPNYEFYKJLVG 85



Db	219	MMHNTSNTSIFRLDNDHITDVKMGLAEFPQVHICD-VYERGTFENEI-EKTIITYNV	276
QY	86	GAGRRCEAPPA-PNLLTLCDRPDIDLRFITIKFOEYSPNLMGEFRSHDYIIATSDGT	144
Db	277	KVEYECRTINADPFRVIALCDPQKLMFFITFRPFTIPQEGLEFLPGNDYFIST-S	334
QY	145	REGIESLOGVCLTRGMYLNVGSPRGCAVPRKPVSEMPMERDRAHSL-----EPG	199
Db	335	KDLYRIRIGRCGTNNMKVFKVCCAPEDNN-KITLALSNSKSYTDYGALNVANIANDS	393
QY	200	KENLPGDPTSNATSRCAEG-----PLPP-----P	223
Db	394	HNVSHGNNTIAIGTNIGNGQIIGGPQSAQIPINPLSGNNNINGIPTTINSIDQFNRI	453
QY	224	SMRVAAGAGGLALLLVAGAGGCMRRRRAPSRSRHPGR-SRGSGSLGAGGG	282
Db	454	IQPIIGNHVGTNAVGTGVGGGGIIL-----TPDHAAGINMLQPRGGI	499
QY	283	MGPREAEPG---ELGIALRG-----GGADPPCPHYEVSGDYH	320
Db	500	NG---AVPGHHIQTGIRLNNVPIQAHVPSHKGNANSININDH---HHYK-----H	547
QY	321	PVTVIVD-----GPPOSPIVITYTSISVLEWPILTITQLFWRKSKSVTTFLEPVQV	373
Db	548	PNBYKNEELVTYNSGAATSDGNIFALWIMILS-----IFP--L	583
QY	374	ITTSCTMSTFSFTTLNPMQACRAQMGSPRIMQCFWMDRIIGTALFVLIIILGRLNM	433
Db	584	LSIQSCHLSY-----NI-----SASIVSTIALDGIHYL	613
QY	434	HQTTL	438
Db	614	IQITL	618
RESULT 38			
ID	ABB65548	standard; Protein; 652 AA.	
AC	ABB65548;		
DT	26-MAR-2002	(first entry)	
DE	Drosophila melanogaster polypeptide SEQ ID NO 23436.		
KM	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.		
OS	Drosophila melanogaster.		
PN	W0200171042-A2.		
PD	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US09231.		
PR	23-MAR-2000; 2000US-191637P.		
XX	11-JUL-2000; 2000US-0614150.		
PA	(PEKE ) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
XX	N-PSDB; ABL09651.		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -		
XX	Disclosure, SEQ ID NO 23436; 21pp + Sequence Listing; English.		

Query	Match	Local Similarity	21.2%	Score 192;	DB 22;	Length 652;
Batches 103;	Conservative 64;	Mismatches 154;	Indels 164;	Gaps 22		
Sequence	652 AA;					
33	VYVNSANKRQEQEG-CGYVLXPQIG-----DRDLDCPRARPPGHSSPNYEFYKYLTVG	85				
Db	219 MHWNTSISRINDTNDTHIIDVKNKMLAEFPQVHICP-VYEPGFENET-EKYLIIYNVS	276				
Qy	86 GAQGRPCGEAPRA-PNLLTCDRPPDDLAEFTTKPQESPNLMGHEFRSHHNYIATSDGR	144				
Db	277 KVEETGRITNADPRVYALCDKPKMLMEFTITFRFTFPQGLSEFLDGNLYFISTIS--S	334				
Qy	145 REGLESLOGGVCLTRGMKYLIRVGQSPRGAVPRKPVSEMERDRGAHSL-----EPG	199				
Db	335 KDLRYRIGGRCSGTNNMKVVFVKCAPEDNN-KTTLASNSKSVTDGALNVNIIANNDES	393				
Qy	200 KENLPGRPTNANSRABE-----PLRP-----P	223				
Db	394 HVNSHGNNIATIGINIGINGQIIGEPQAGIPINDLSGNNNINGIPTTINSINDQFRAP	453				
Qy	224 SMPVAGAAAGLALLLLVAGAGAGAMCWRRBRAPXSESRHPFG-SFGRGSGSLGGGGG	282				
Db	454 IQPMIIGHNHTNAVGTGIVGGGILL-----TPGHAGNINMLQPRGGGI	499				
Qy	283 MGPRAEPFG---ELGIALRG-----GGAADPPRCRYEKVSGDYGH	320				
Db	500 NG--AYPGHHHITGTIRINNVPOTHNYPSSHKNANSNININDH--HHYNK--H	547				
Qy	321 PVTIVOD-----GPQSPNIYYTISIVLEPMILHTIQGLFPMRSKCSRVTFLFPVQV	373				
Db	548 PNEVVKAEELTYSAGATSDGNIIPALMIWLS-----IFP--L	583				
Qy	374 ITTSTCWTSFSEFTTLNPSMQACRAQMGSEFRIMCFWGDRIILGALVVLVLIILGLNLM	433				
Db	584 LSIQSCHLSSY-----WI-----SASELVSTIALIGIHYL	613				
Qy	434 HQTTL 438					
Db	614 IQITL 618					
RESULT 39						
AA882605						
ID	AA882605 standard; Protein; 234 AA.					
XX	AA882605;					
AC						
XX	16-MAY-1996 (first entry)					
DT						
XX						
XX	Eph transmembrane tyrosine kinase family ligand, Efl-2.					
DE						
XX	Efl-2, EHK1-L; Eph transmembrane tyrosine kinase family ligand;					
KM	neurological disorder; identification; diagnosis.					
XX						
OS	Homo sapiens.					
XX						
FH	Key	Location/Qualifiers				
FT	Peptide	1..30				
FT		/label= signal_peptide				

```

FT Region 218..235
PT /note= "carboxy terminal hydrophobic
FT Misc-difference 158 GPI-recognition tail"
FT /note= "residue borders main conserved regions"
XX WO9527060-A2.
XX 12-OCT-1995.
XX 04-APR-1995; 95WO-US04208.
XX 21-OCT-1994; 94US-0327423.
XX 04-APR-1994; 94US-0228075.
XX 12-APR-1994; 94US-0228402.
XX 01-SEP-1994; 94US-0299567.
XX (REGG-) REGENERON PHARM INC.
XX Aldrich TM, Davis S, Gale N, Goldfarb M, Maisonnier PC;
XX Yancopoulos GD;
XX WPI; 1995-358635/46.
XX N-PSDB; AA03883.
XX Ligands which bind Eph family receptors - used in the diagnosis of
XX neurological disorders
XX Disclosure; Fig 2; 58pp; English.
XX Efl-2 (also known as EHK-11) is an Eph transmembrane tyrosine kinase
XX family ligand. It has homology with B61 (Efl-1) (see AA82604). Efl-2
XX ends in a C-terminal hydrophobic sequence that appears to be a
XX recognition sequence allowing it to be GPI-linked and thus lacking in
XX an intracellular domain. Efl-2 is useful for identifying other ligands
XX for EHK-1, -2, -3, Bck and Elk receptors. The ligands are useful in
XX promoting a differential function and/or influencing the phenotype,
XX such as growth and/or proliferation, of receptor bearing cells. They
XX may be used in the diagnosis, and treatment of neurological disorders.
XX Sequence 234 AA;
SQ
Query Match 7.4%; Score 181; DB 16; Length 234;
Best Local Similarity 28.4%; Pred. No. 3.66-07;
Matches 64; Conservative 23; Mismatches 82; Indels 56; Gaps 11;
QY 7 GPGGVRVGAALLLLGLVLSGLSEPVYVNSANKRFOAEGGVLYPOIGRDLILCP--R 64
DB 24 GPG-----GALG-----NRHAVYVNSNGHRLRE-GTVQVNVNDYLDIYCPHYN 67
QY 65 ARPPGHSHPNYEFYKLYLVGAQGRCEAPAPNLLTCDPDL---DLRTTIKFOEYS 121
DB 68 SSGAGGPGGAGGEOVYLVVWSNNGYRTCAAGGFR-RWECNRPHAPHSPIRSEKFOYS 126
QY 122 PVLWGHFHSHDYIITATSDGTREGLESLOGGVCLTGMKYLRLVVGSPGGAVPKREV 181
DB 127 APLSGEFLHAGHEHYIISTPTNHL-----WKCLR--MKVFCVCASTSHSG--EKPV 174
QY 182 SEMP-----MERDRGAHSL-----PKGNLP 204
DB 175 PVLPGFTMGPNKINVLDFEGENPOVKLEKISGTSFKREHLP 219
RESULT 40
AA071481
ID AA071481 standard; Protein; 238 AA.
XX AA071481;
AC
XX 03-OCT-1995 (first entry)
DT
XX
XX Human hek-L protein.
DE
XX

```

```

XX Ligand; cell surface; tyrosine kinase receptor; tumorigenesis;
XX Immunogen.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..19
XX Protein /note= "signal peptide"
XX 20..238
XX WO9506065-A.
XX 02-MAR-1995.
XX 17-AUG-1994; 94WO-US09282.
XX 20-AUG-1993; 93US-0109745.
XX 30-AUG-1993; 93US-0114426.
XX 03-DEC-1993; 93US-0161132.
XX 09-MAY-1994; 94US-0240124.
XX (IMMEX) IMMUNEX CORP.
XX Beckmann MP, Cerretti DP;
XX WPI; 1995-106811/14.
XX N-PSDB; AA085887.
XX New isolated DNA encoding hek-L protein or its fusion products -
XX useful as assay reagent or for carrying therapeutic and
XX diagnostic compounds to leukaemia cells.
XX Claim 21; Page 36; 45pp; English.
XX The sequence is that of a novel protein designated hek-L, a protein
XX that can bind hek (a cell surface receptor tyrosine kinase). Hek-L
XX is the first known ligand for hek and can be used to study cellular
XX processes regulated by hek (which may be involved in tumorigenesis).
XX It is also an immunogen for antibody production, as a reagent for
XX detecting hek or hek-L in in vitro assays, to determine binding of
XX hek proteins, to purify hek proteins, and to carry diagnostic or
XX cytotoxic agents to particular leukaemia cells that express the hek
XX antigen. Hek-L also binds the elk tyrosine kinase receptors.
XX See also AA071482.
XX Sequence 238 AA;
SQ
Query Match 7.3%; Score 179; DB 16; Length 238;
Best Local Similarity 28.4%; Pred. No. 5.46-07;
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;
QY 7 GPGGVRVGAALLLLGLVLSGLSEPVYVNSANKRFOAEGGVLYPOIGRDLILCP--R 64
DB 24 GPG-----GALG-----NRHAVYVNSNGHRLRE-GTVQVNVNDYLDIYCPHYN 67
QY 65 ARPPGHSHP---NYEFYKLYLVGAQGRCEAPAPNLLTCDPDL---DLRTTIK 117
DB 68 SSGAGGPGGAGGEOVYLVVWSNNGYRTCAAGGFR-RWECNRPHAPHSPIRSEK 126
QY 118 QEYSPVLWGHFHSHDYIITATSDGTREGLESLOGGVCLTGMKYLRLVVGSPGGA 177
DB 127 QRYSAFSLGIEFHAGHEHYIISTPTNHL-----WKCLR--MKVFCVCASTSH 174
QY 178 RKPVSSEMP-----MERDRGAHSL-----PKGNLP 204
DB 175 EKVPFTLPQFTMGPNKINVLDFEGENPOVKLEKISGTSFKREHLP 223
RESULT 41
AA002587
ID AA002587 standard; Protein; 184 AA.
XX AA002587;
AC

```

```

XX 28-NOV-1996 (first entry)
DT
XX Lerk-6 protein.
DE
XX Lerk-6; hek; elk; cell surface receptor; culture; reagent;
KM neuron; disorder; injury; delivery agent; diagnostic; therapeutic;
XX Lerk-7; probe; cytokine.
XX Mus musculus.
OS
XX MO9617925-A1.
PN
XX 13-JUN-1996.
PD
XX 05-DEC-1995; 95WO-US15781.
PF
XX 01-MAR-1995; 95US-0396946.
PR
XX 06-DEC-1994; 94US-0351025.
XX
XX (IMMV) IMMUNEX CORP.
PA
XX Cerretti DP;
PI
XX WPI; 1996-287171/29.
DR
XX N-PSDB; AAT32700.
XX
XX New isolated human Lerk-7 cytokine - which binds to cell surface
PT receptor elk, hek and eck, useful for delivering agents to cells or
XX for treating neural disorders
XX
XX Example 1; Page 35; 49p; English.
XX
XX The present sequence is that of a murine cytokine, Lerk-6, encoded by
CC AAT32700. A fragment of murine Lerk-6 DNA was isolated by PCR for use as
CC a probe for a human Lerk-6 homologue. The probe however led to the
CC identification of a human Lerk-7 gene (see AAT3659). The Lerk-7 gene
CC and protein can be used for studying the role of Lerk-7 in conjunction
CC with elk, hek and eck receptors. They can also be used for delivering
CC diagnostic or therapeutic agents to cells, e.g. cancer cells. The
CC Lerk-7 proteins can also exhibit neuroprotective or neurotrophic
CC properties and can be used to treat neural tissue disorders.
XX
XX Sequence 184 AA;
SQ
Query Match 7.2%; Score 176; DB 17; Length 184;
Best Local Similarity 29.3%; Pred. No. 7e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;
QY 33 VYVNSANKRFQAE-----GGYVLYPQIGDRDLCPAPRPGPSNVEFYKLYLVGA 87
DB 10 VYVNSRNPFRQVSAVGDGGYTVESINDYLDICPHYGAPLP-PARMERKYLIMVNGS 68
QY 88 QGRRCAPAPNLLLTCDRPLD--DLRFTIKQEVSPNIMGHEFRSHHDYIATSDGT 144
DB 69 GHASCDHRQGRGFKWECNRPAPAGPLKFSKPOLFTPFSIGFEFRDQHEYYIATP-- 126
QY 145 REGLESLOGGVCLTRGKVLIRVGQSPRGAVPRKPVSEKMERDRGAHSLBEGKENTP 204
DB 127 ----PNIYDRPCLR-LKYYR-----PTNETLY 149
QY 205 GDP-----TSNATSRGAG 218
DB 150 EAPFPIFTSNSSCSGLG 167

```

```

RESULT 42
AAR91283
ID AAR91283 standard; Protein; 184 AA.
XX
XX AAR91283;
AC
XX 10-OCT-1996 (first entry)
DT

```

```

XX LERK-6 protein.
DE
XX LERK-6; hek; elk; cell surface receptors; culture; reagent;
KM neurons; disorder; injury; delivery agent; diagnostic; therapeutic.
XX Mus musculus.
OS
XX WO9610911-A1.
PN
XX 18-APR-1996.
PD
XX 04-OCT-1995; 95WO-US12779.
PF
XX 03-OCT-1995; 95US-0538709.
PR
XX 05-OCT-1994; 94US-0318393.
XX
XX (IMMV) IMMUNEX CORP.
PA
XX Cerretti DP;
PI
XX WPI; 1996-209575/21.
DR
XX N-PSDB; AAT14009.
XX
XX Isolated DNA encoding cytokine designated LERK-6 which binds to hek
PT and elk cell surface receptors - useful for drug delivery and
XX screening procedures.
XX
XX Claim 6; Page 34; 44p; English.
XX
XX The LERK-6 polypeptide encoded can be used to isolate cells
CC expressing hek/elk cell surface receptors, or to measure the
CC biological activity of such receptors. The protein may also be used
CC as a delivery agent, taking diagnostic and therapeutic agents to
CC cells expressing such receptors. LERK-6 can also be used as a tissue
CC culture reagent to enhance the viability or prolong the lifespan of
CC the neurons. Neural tissue disorders and injuries may be treated by
XX contact with the polypeptide.
XX
XX Sequence 184 AA;
SQ
Query Match 7.2%; Score 176; DB 17; Length 184;
Best Local Similarity 29.3%; Pred. No. 7e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;
QY 33 VYVNSANKRFQAE-----GGYVLYPQIGDRDLCPAPRPGPSNVEFYKLYLVGA 87
DB 10 VYVNSRNPFRQVSAVGDGGYTVESINDYLDICPHYGAPLP-PARMERKYLIMVNGS 68
QY 88 QGRRCAPAPNLLLTCDRPLD--DLRFTIKQEVSPNIMGHEFRSHHDYIATSDGT 144
DB 69 GHASCDHRQGRGFKWECNRPAPAGPLKFSKPOLFTPFSIGFEFRDQHEYYIATP-- 126
QY 145 REGLESLOGGVCLTRGKVLIRVGQSPRGAVPRKPVSEKMERDRGAHSLBEGKENTP 204
DB 127 ----PNIYDRPCLR-LKYYR-----PTNETLY 149
QY 205 GDP-----TSNATSRGAG 218
DB 150 EAPFPIFTSNSSCSGLG 167

```

```

RESULT 43
AAV06820
ID AAV06820 standard; Protein; 184 AA.
XX
XX AAV06820;
AC
XX 24-JUN-1999 (first entry)
DT
XX Murine LERK-6 polypeptide.
DE
XX LERK-6 polypeptide; hek receptor; elk receptor; human; murine; injury;
KM

```

KW cell proliferation; neural growth; neural tissue; neurological disease;  
 KW neurodegenerative; excitotoxicity.  
 OS Mus sp.  
 PN WO9910495-A1.  
 XX  
 XX 04-MAR-1999.  
 PD  
 XX 27-AUG-1998; 98WO-US17772.  
 PF  
 XX 29-AUG-1997; 97US-0920440.  
 PR  
 XX (IMMUNEX CORP.  
 PA  
 XX Cerretti DP;  
 PI  
 XX WPI; 1999-243567/20.  
 DR  
 XX N-PSDB; AAX32761.  
 PT  
 XX New cytokine designated LERK-6  
 PS  
 XX Claim 6; Page 37-38; 46pp; English.  
 XX  
 CC The invention relates murine and human LERK-6 polypeptides that bind to  
 CC hek/elk receptors. Host cells transfected or transformed with vectors  
 CC comprising the LERK-6 nucleic acid sequences are used for the recombinant  
 CC production of the proteins. LERK-6 polypeptides may be useful in the  
 CC enhancement, stimulation, proliferation or growth of cells expressing the  
 CC hek or elk receptor. The ligand and receptor complex may be involved in  
 CC neural growth, development and/or maintenance. LERK-6 can be used for  
 CC treating disorders of neural tissue such as injury or neurological  
 CC diseases, either chronic or acute. LERK-6 may be employed in treating  
 CC neurodegenerative conditions where there is neural death, excitotoxicity.  
 CC In addition, they may be administered to a mammal to exert a trophic  
 CC effect on neural tissue. They can also be used as reagents for those  
 CC conducting quality assurance studies e.g. to monitor shelf life and  
 CC stability of elk protein under different conditions. The polypeptides can  
 CC also be used as carriers for delivering agents attached to cells bearing  
 CC the elk or hek cell surface receptor. The present sequence represents a  
 CC murine LERK-6 polypeptide.  
 CC  
 XX  
 SQ Sequence 184 AA;  
 Query Match 7.2%; Score 176; DB 20; Length 184;  
 Best Local Similarity 29.3%; Pred. No. 7e-07;  
 Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;  
 QY 33 VYVNSANKRFOAE-----GGYVLYPQIGRLDLCPRARPPGHSSPNVEFYKLYVGGA 87  
 DB 10 VYVNSNRPFCVSAVGDGGYVEVSINDYDIYCPHYGAPLP-PAERMYRYILYVNGE 68  
 QY 88 QGRRCAPAPNLLTCDRPLD---DLRFTIKFOEYSPNLMWGHFRSHDYIATSDGT 144  
 DB 69 GHASCDHRQGRKMECNRPAPGGLKFKSEKQLFTPSLGFEFRPHGHYIYISATP-- 126  
 QY 145 REGLSLGGVCLTRGMKYLRLVVGSPRGAVPRKVSSEMPERDRGAHSLPEKENTLP 204  
 DB 127 ----PNLVDRPCLR--LKYYVR-----PTNETLY 149  
 QY 205 GDP-----TSNATSRGAEG 218  
 DB 150 EAPPEFTSNSSCSGLQG 167  
 RESULT 44  
 AAR94766  
 ID AAR94766 standard; Protein; 209 AA.  
 AC AAR94766;  
 XX  
 XX 02-JUL-1996 (first entry)  
 DT  
 XX

DE Mouse EPH receptor ligand Elf-1.  
 XX  
 XX Elf-1; EPH receptor ligand; dementia; tachycardia; therapy;  
 KW diagnosis; transgenic animal.  
 KW  
 OS Mus musculus.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT /label= Sig\_peptide  
 FT Protein 21..209  
 FT /label= Mat\_protein  
 FT Region 69..159  
 FT /label= Cys4\_motif  
 FT Region 35..166  
 FT /label= Core\_sequence\_motif  
 FT Modified-site 38  
 FT /note= "potential N-glycosylation site"  
 FT Modified-site 170  
 FT /note= "potential N-glycosylation site"  
 FT Modified-site 184  
 FT /note= "potential N-glycosylation site"  
 FT  
 XX  
 XX WO9609384-A1.  
 XX  
 XX 28-MAR-1996.  
 XX  
 XX 19-SEP-1995; 95WO-US11869.  
 XX  
 XX 27-FEB-1995; 95US-0393462.  
 XX 19-SEP-1994; 94US-0308814.  
 XX  
 XX (HARD ) HARVARD COLLEGE.  
 PA  
 XX  
 PI Cheng H, Flanagan UG;  
 XX  
 XX WPI; 1996-188446/19.  
 DR N-PSDB; AAT15008.  
 XX  
 PT Murine and chicken EPH receptor ligand, Elf-1 - useful in diagnosis  
 PT and treatment of disorders associated with the Elf-1 gene, e.g.  
 PT dementia, tachycardia, etc.  
 PT  
 XX  
 XX  
 PS Claim 1; Page 86; 107pp; English.  
 CC A novel mouse EPH receptor ligand, Elf-1 (AAR94766), is involved in  
 CC the formation and maintenance of ordered spatial arrangements of  
 CC differentiated tissue. It is the product of a cDNA clone (AAT15008)  
 CC obtd. from an embryo mid- and hind-brain cDNA expression library.  
 CC This cDNA is used for the prodn. of recombinant Elf-1, which can  
 CC be used to modulate proliferation, survival and/or differentiation  
 CC of cells and tissues, and to stimulate or antagonise intracellular  
 CC signal transduction pathways mediated by the EPH-type receptor.  
 CC  
 XX  
 SQ Sequence 209 AA;  
 Query Match 7.2%; Score 176; DB 17; Length 209;  
 Best Local Similarity 29.3%; Pred. No. 8.2e-07;  
 Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;  
 QY 33 VYVNSANKRFOAE-----GGYVLYPQIGRLDLCPRARPPGHSSPNVEFYKLYVGGA 87  
 DB 35 VYVNSNRPFCVSAVGDGGYVEVSINDYDIYCPHYGAPLP-PAERMYRYILYVNGE 93  
 QY 88 QGRRCAPAPNLLTCDRPLD---DLRFTIKFOEYSPNLMWGHFRSHDYIATSDGT 144  
 DB 94 GHASCDHRQGRKMECNRPAPGGLKFKSEKQLFTPSLGFEFRPHGHYIYISATP-- 151  
 QY 145 REGLSLGGVCLTRGMKYLRLVVGSPRGAVPRKVSSEMPERDRGAHSLPEKENTLP 204  
 DB 152 ----PNLVDRPCLR--LKYYVR-----PTNETLY 174  
 QY 205 GDP-----TSNATSRGAEG 218

Db 175 EAPEPFTSNSSCSGLG 192

## RESULT 45

AAW71006  
ID AAW71006 standard; Protein; 209 AA.  
XX  
AC AAW71006;  
XX  
DT 20-OCT-1998 (first entry)  
XX  
DE Amino acid sequence of a mammalian E1f-1 protein.  
XX  
KW Mouse; EPH receptor ligand; E1f-1; mek-4; sek-Ap; tyrosine kinase ligand;  
KW B61; LERK-2; proliferation; differentiation; intracellular signalling;  
KW increased survival; neuronal cell; neuron survival; treatment;  
KW Alzheimer's; Parkinson's; lymphatic tumour; artificial liver; cartilage;  
KW bone formation.

XX Mus sp.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..20  
FT /note= "signal peptide"  
FT Protein 21..209  
FT /note= "mature protein"  
FT Modified-site 38  
FT /note= "potential N-linked glycosylation site"  
FT Misc-difference 35..166  
FT /note= "contains core sequence motif"  
FT Misc-difference 69..159  
FT /note= "contains a Cys4 motif"  
FT Modified-site 170  
FT /note= "potential N-linked glycosylation site"  
FT Modified-site 184  
FT /note= "potential N-linked glycosylation site"

XX US5795734-A.

XX 18-AUG-1998.

XX 31-MAY-1995; 95US-0455001.

XX 31-MAY-1995; 95US-0455001.

XX 19-SEP-1994; 94US-0308814.

XX 27-FEB-1995; 95US-0393462.

XX (HARD ) HARVARD COLLEGE.

XX Cheng H, Flanagan JG;

XX WPI; 1998-466665/40.

XX N-PSDB; AAV42926.

XX Nucleic acid encoding E1f-1 protein that binds to EPH-type receptor  
FT - for production of E1f-1 protein, useful for regulating  
FT proliferation, differentiation, and survival of cells  
XX  
XX Claim 1; Fig 2A; 53pp; English.

XX The present sequence represents a mammalian EPH receptor ligand  
CC designated E1f-1. This ligand can bind to both mek-4 and sek-Ap. E1f-1 is  
CC a tyrosine kinase ligand, which is linked to the membrane through a  
CC phosphatidylinositol linkage. It shares some homology to 2 other EPH  
CC receptor ligands, B61 and LERK-2. The E1f-1 protein modulates  
CC proliferation, differentiation and survival of EPH receptor-expressing  
CC cells by stimulating or antagonising intracellular signalling mediated  
CC by the EPH receptor. Typical of many potential applications are  
CC increasing survival of neuronal cells in culture (e.g. where intended for  
CC transplantation), also therapeutically in increase neuron survival  
CC (e.g. treatment of Alzheimer's or Parkinson's diseases), to prevent  
CC nervous system and lymphatic tumours, to induce differentiation of

CC hepatocytes to form an artificial liver, to induce cartilage and bone  
CC formation.

XX Sequence 209 AA;

Query Match 7.2%; Score 176; DB 19; Length 209;  
Best Local Similarity 29.3%; Pred. No. 8.2e-07;  
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 YVMSANKRPOAE-----GGVLYPQIGDRLDLCRRAPRPPGPHSPVYEFKYLVGGA 87  
DB 35 YVMSNSRPFQVSAVGDGGGIVTVSINDYDIPGHGAPLP-PAEMKEXILYLVNAGE 93  
QY 88 QGRRCAPAPNLLITCDRPLD---DLRFITKQFYSNVLWGHFRSHDYIATSDGT 144  
DB 94 GNASCDHRQGRGKRWCKRCPAPRPGPLKFSEKFLQFTFSLGFEPRPGHEVYISATP-- 151  
QY 145 RGLSLGCGVCLTGMVYLKRVGSPRGAVPRKRVSEMEMERDGAHLEPKENLP 204  
DB 152 ---PNLVDRPCLR--LKVVYR-----PYNRETLV 174  
QY 205 GDP---TSNATSRGAEG 218  
DB 175 EAPEPFTSNSSCSGLG 192

## RESULT 46

AAW06822  
ID AAW06822 standard; Protein; 213 AA.

XX AAW06822;

XX 24-JUN-1999 (first entry)

XX Human LERK-6 polypeptide.

XX LERK-6 polypeptide; hek receptor; elk receptor; human; murine; injury;  
XX cell proliferation; neural growth; neural tissue; neurological disease;  
XX neurodegenerative; excitotoxicity.

XX Homo sapiens.

XX WO9910495-A1.

XX 04-MAR-1999.

XX 27-AUG-1998; 98WO-US17772.

XX 29-AUG-1997; 97US-0920440.

XX (IMMUNEX ) IMMUNEX CORP.

XX Cerretti DP;

XX WPI; 1999-243567/20.

XX N-PSDB; AAX32767.

XX New cytokine designated LERK-6

XX Claim 6; Page 42; 46pp; English.

XX The invention relates murine and human LERK-6 polypeptides that bind to  
CC hek/elk receptors. Host cells transfected or transformed with vectors  
CC comprising the LERK-6 nucleic acid sequences are used for the recombinant  
CC production of the proteins. LERK-6 polypeptides may be useful in the  
CC enhancement, stimulation, proliferation or growth of cells expressing the  
CC hek or elk receptor. The ligand and receptor complex may be involved in  
CC neural growth, development and/or maintenance. LERK-6 can be used for  
CC treating disorders of neural tissue such as injury or neurological  
CC diseases, either chronic or acute. LERK-6 may be employed in treating  
CC neurodegenerative conditions where there is neural death, excitotoxicity.  
CC In addition, they may be administered to a mammal to exert a trophic  
CC effect on neural tissue. They can also be used as reagents for those

CC conducting quality assurance studies e.g. to monitor shelf life and  
 CC stability of elk protein under different conditions. The polypeptides can  
 CC also be used as carriers for delivering agents attached to cells bearing  
 CC the elk or hek cell surface receptor. The present sequence represents a  
 CC human LER-6 polypeptide.

XX Sequence 213 AA;

Query Match 7.2%; Score 175.5; DB 20; Length 213;  
 Best Local Similarity 36.8%; Pred. No. 9.3e-07;  
 Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;

QY 33 VYVNSANKRFOA-----EGGYLYTPIQIGRLDLCPRARPSPHSPPYEPYKLYVGA 87  
 Db 39 VYVNSNPFHAGAGDDGGGYTVEVSINDYDICYPHYGAPLP-PAERKEHYVLYMNGE 97  
 QY 88 QGRCEAPAPNLLTCDRPLD--DLRFTIKFOEYSPNLMGHEFRSHDYIATIS 141  
 Db 98 GHSACDHRCGRKRCWECNRPAAPGAPLKFSEKQULFTFSLGSEFRPCHERYIISAT 154

RESULT 47

AAB54058  
 ID AAB54058 standard; Protein; 218 AA.

AC AAB54058;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:510.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KM neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KM antinflammatory; cardiant; gene therapy; chromosome mapping;  
 KM linkage analysis; tissue identification; tissue typing; forensic;  
 KW neural; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative.

OS Homo sapiens.

XX WO200055320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05989.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-579444/54.

XX N-PSDB; AAC98623.

PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX  
 PS Claim 11; Page 946-947; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, cardiant and antinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.

CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.

XX Sequence 218 AA;

Query Match 7.0%; Score 172.5; DB 21; Length 218;  
 Best Local Similarity 28.3%; Pred. No. 1.7e-06;  
 Matches 51; Conservative 30; Mismatches 80; Indels 19; Gaps 6;

QY 8 PGVVRGALLLGLVGLVGLS---EPYVNSANKRFOAEGGYLYTPIQIGRLDLCR 64  
 Db 8 PGQARAMEFLMAPDLGLCCSLAADRHIVFWSSNPKFKNV-DYIIHVLNDYDLCFH 66  
 QY 65 ARPPGPHSSFN--YEFYKLYVGAQGRCEAPAPNLLTCDRPLD--DLRFTIKFOE 119  
 Db 67 YE---DHSVADAMEGYIILYVHEHEYQLCQPSKQVWQCNRPKAKHGPELSEKFOR 123  
 QY 120 YSPNLMGHEFRSHDYIATSDGTREGLESLOGGVCLTRGKMYLIRVGQSRGCAVPPK 179  
 Db 124 FTFTLGKEFKESHYYIYSKPIHEDR-----CLRLKTVSGKITHSPOAHNDPGE 176

RESULT 48

AAR94767  
 ID AAR94767 standard; Protein; 200 AA.

XX AAR94767;

DT 02-JUL-1996 (first entry)

DE Chicken EPH receptor ligand Elf-1.

XX Elf-1; EPH receptor ligand; dementia; tachycardia; therapy;  
 KW diagnosis; transgenic animal.

XX Gallus sp.

XX Key

FT Reptide 1..21 Location/Qualifiers

FT Protein 22..200 /label= Sig\_peptide

FT Region 61..150 /label= Mat\_protein

FT Region 33..157 /label= Cys4\_motif

FT Region /label= Core\_sequence\_motif

XX WO9609384-A1.

XX 28-MAR-1996.

XX 19-SEP-1995; 95WO-US11869.

XX 27-FEB-1995; 95US-0393462.

XX 19-SEP-1994; 94US-0308814.

XX (HARD ) HARVARD COLLEGE.

XX Cheng H, Flanagan DG;

XX WPI; 1996-188446/19.  
 DR N-PSDB; AAT15009.  
 PT Murine and chicken EPH receptor ligand, Elf-1 - useful in diagnosis

PT		and treatment of disorders associated with the E1f-1 gene, e.g.
Pt		dementia, tachycardia , etc.
XX		
PS	Claim 1, Page 88-89, 107pp; English.	
XX		
CC	A novel chicken Eph receptor ligand, E1f-1 (AA94767), is involved in	
CC	the formation and maintenance of ordered spatial arrangements of	
CC	differentiated tissue. It is the product of a cDNA clone (AA15009)	
CC	obt'd from an embryo cDNA expression library. This cDNA is used for	
CC	the prodn. of recombinant E1f-1, which can be used to modulate	
CC	proliferation, survival and/or differentiation of cells and tissues,	
CC	and to stimulate or antagonise intracellular signal transduction	
CC	pathways mediated by the Eph-type receptor.	
XX		
SQ	Sequence      200 AA;	
	Query Match            7.0%; Score 172; DB 17; Length 200;	
	Best Local Similarity   35.3%; Pred.No. 1.7e-06;	
	Matches     49; Conservative   16; Mismatches   58; Indels   16; Gaps	
Qy	15 AALLGLVGVSGHSLP-----VYMNANRRFOAGGYLYLPQIGRDLCPRA 65	
Db	7 AALLAALVG-VCMWSDDPKGVISDRYAIVMNSNRFH-RDITYEVSINDYDIYCPH 64	
	:           :           :           :	
Qy	66 RPPGPHSSPNVEFYKLVLVGAGGRRCAPAPNLILTCDDPD--DLRAFTIKQEVSP 122	
Db	65 EEPPLP-AEMERERVLYMVNYEGSHACDRQRKFPRMECNRPDSPGCLTFSEKFLFTP 122	
	:           :           :           :	
Qy	123 NLNGHEFRSHDYIIATIS 141	
Db	123 FSLGFERRPGHEHYIYSAS 141	
	:           :           :           :	
RESULT 49		
ID	AAW71007 standard; Protein; 200 AA.	
AAW71007		
AC	AAW71007;	
DT	20-OCT-1998 (first entry)	
XX		
DE	Amino acid sequence of an avian E1f-1 protein.	
XX		
KW	Chicken; Eph receptor ligand; E1f-1; mek-4; sek-AP;	
KW	Cytosine kinase ligand; B61; LERK-2; proliferation; differentiation;	
KW	intracellular signaling; increased; survival; neuronal cell;	
KW	neuron survival; treatment; Alzheimer's; Parkinson's; lymphatic tumour;	
KW	artificial liver; cartilage; bone formation.	
OS	Gallus sp.	
XX		
Key	Location/Qualifiers	
FH	Misc-difference 61..150	
FT	/note= "contains a Cys4 motif"	
FT	Misc-difference 35..157	
FT	/note= "contains a core sequence motif"	
XX		
FN	US5795734-A.	
XX		
PD	18-AUG-1998.	
XX		
PF	31-MAY-1995; 95US-0455001.	
PR	31-MAY-1995; 95US-0455001.	
PR	19-SEP-1994; 94US-0308614.	
PR	27-FEB-1995; 95US-0393462.	
XX		
PA	(HARD ) HARVARD COLLEGE.	
XX		
PI	Cheng H, Flanagan JG;	
XX		
WP	WPI, 1999-466665/40.	
DR	N-PSTD; AAV42927.	

XX	NCIcic acid encoding E1f-1 protein that binds to Eph-type receptor
PT	- for production of E1f-1 protein, useful for regulating
PT	proliferation, differentiation, and survival of cells
XX	
PS	Claim 1; Columns 75-78; 53pp; English.
XX	
CC	The present sequence represents an avian Eph receptor ligand designated
CC	E1f-1. This ligand can bind to both mek-4 and sek-AP. E1f-1 is a
CC	tyrosine kinase ligand, which is linked to the membrane through a
CC	phosphatidylinositol linkage. It shares some homology to 2 other Eph
CC	receptor ligands, B61 and LERK-2. The E1f-1 protein modulates
CC	proliferation, differentiation and survival of Eph receptor-expressing
CC	cells by stimulating or antagonising intracellular signalling mediated
CC	by the Eph receptor. Typical of many potential applications are
CC	increasing survival of neuronal cells in culture (e.g. where intended for
CC	transplantation), also therapeutically in increase neuron survival
CC	(e.g. treatment of Alzheimer's or Parkinson's disease), to prevent
CC	nervous system and lymphatic tumours, to induce differentiation of
CC	hepatocytes to form an artificial liver, to induce cartilage and bone
CC	formation.
XX	
SQ	Sequence 200 AA:
	Query Match 7.0%; Score 172; DB 19; Length 200;
	Best Local Similarity 35.3%; Pred. No. 1.7e-06;
	Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5
QY	15 ALTLGLVGLVSGLSLEP-----VYVNSANKRFGAEGGYLYLPQIGRLDILLCPRA 65
DB	7 AALLAAIVG-VCTWSDPDGKVTSDRAVYVWNSNRFH-RGDIYVENSINDYLDIYCPHY 64
QY	66 RPPGPHSSPVYEFYKLYLVGAQGRRCCEAPAPNLLITCDRPLL---DLRPTIKFGQYSP 122
DB	65 EEPLEP-AEENKERYVLVYVNYEGHASCDRQGFGRWECNRPDPSGPKRSEKFLQFLTP 122
QY	123 NUNGHFRSHHDYIYIATTS 141
DB	123 FSLGFEFRPGHEHYIYSAS 141
RESULT 50	
AAAR71482	
ID	AAAR71482 standard; Protein; 201 AA.
XX	
AC	AAAR71482;
XX	
DT	03-OCT-1995 (first entry)
XX	
DE	Human hek-L protein.
XX	
KW	Ligand; cell surface; tyrosine kinase receptor; tumorigenesis;
XX	immunogen.
XX	
OS	Homo sapiens.
XX	
FT	Key Location/Qualifiers
FT	Peptide 1..22
FT	/note= "signal peptide"
FT	Protein 23..201
XX	
PN	WC9506065-A.
XX	
PD	02-MAR-1995.
XX	
PF	17-AUG-1994; 94MO-US09282.
XX	
PR	20-AUG-1993; 93US-0109745.
PR	30-AUG-1993; 93US-0114426.
PR	03-DEC-1993; 93US-0161132.
PR	09-MAY-1994; 94US-0240124.
XX	
PA	(IMMV ) IMMUNEX CORP.



XX Beckmann MP, Cerretti DP;  
 PI WPI, 1995-106811/14.  
 DR N-PSDB; AAQ8588.  
 XX  
 XX New isolated DNA encoding hek-L protein or its fusion products -  
 PI useful as assay reagent or for carrying therapeutic and  
 PI diagnostic compounds to leukaemia cells.  
 XX  
 XX Claim 21; Page 38; 45pp; English.  
 XX  
 CC The sequence is that of a novel protein designated hek-L, a protein  
 CC that can bind hek (a cell surface receptor tyrosine kinase). Hek-L  
 CC is the first known ligand for hek and can be used to study cellular  
 CC processes regulated by hek (which may be involved in tumorigenesis).  
 CC It is also an immunogen for antibody production, as a reagent for  
 CC detecting hek or hek-L in in vitro assays, to determine binding of  
 CC hek proteins, to purify hek proteins, and to carry diagnostic or  
 CC cytotoxic agents to particular leukaemia cells that express the hek  
 CC antigen. Hek-L also binds the elk tyrosine kinase receptors.  
 CC See also AAR71481.

XX  
 SQ Sequence 201 AA;

Query Match 6.9%; Score 169.5; DB 16; Length 201;  
 Best Local Similarity 29.9%; Pred. No. 2.8e-06;

Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGSL-EPVYWNANKRFQEGGVLYPQIGRLDLCPRARPPGHSSPNHFFKLY 82  
 DB 20 LRGGSLRHRVYWNSSNPRL-LRGDAVELGLNDYLDIVCPHYEGPPEGP--ETPALY 76  
 QY 83 LVGAQGRCEAP-PAPNLLTCDRPDLRFPTIKOEKSPNLWGHEFRSHDYIATS 141  
 DB 77 MYDMPEYECQABGPAYKWCSPFGHVQSEKIQKPTPSLGFELPGETTYIISVP 136  
 QY 142 DGTREGLSLQGVCLTRGMKVLNVGSPRGAVPRKRVSEMPENRDRGAHSLPEGKE 201  
 DB 137 --TPE-----SSGQCL-----RLQVSVCKEKESASHPV----- 164  
 QY 202 NLPDPTSNATS--RGABGLPFPMPNPAVAGAGLALLL 240  
 DB 165 ----GSPGESGTSGRKGGDTTSP-----LCILL 189

Search completed: February 11, 2003, 12:02:56  
 Job time : 43.7736 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 12:01:50 / Search time 14.3082 Seconds  
(without alignments)  
935.649 Million cell updates/sec

Title: US-10-021-121-2

Perfect score: 2450

Sequence: 1 MGPHSGPGVGRVALLLG.....TTLRQASVEMAGGHP 455

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 50 summaries

Database :

Issued Patents: AA:\*

- 1: /cgn2\_6/ptodata/1/aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/aa/PCTus.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1837	75.0	340	US-09-214-631-3	Sequence 3, Appl1
2	628.5	25.7	336	US-08-436-044-2	Sequence 2, Appl1
3	628.5	25.7	336	US-08-436-054-2	Sequence 2, Appl1
4	628.5	25.7	336	PCT-US95-08812-2	Sequence 2, Appl1
5	623	25.4	346	US-08-213-403-2	Sequence 2, Appl1
6	623	25.4	346	US-08-458-077-2	Sequence 2, Appl1
7	623	25.4	346	US-08-460-741-2	Sequence 2, Appl1
8	623	25.4	346	US-08-747-240-2	Sequence 2, Appl1
9	623	25.4	346	US-08-299-557-6	Sequence 6, Appl1
10	620.5	25.3	333	US-08-436-044-4	Sequence 4, Appl1
11	620.5	25.3	333	US-08-436-054-4	Sequence 4, Appl1
12	620.5	25.3	333	US-08-271-948-2	Sequence 2, Appl1
13	620.5	25.3	333	PCT-US95-08534-2	Sequence 2, Appl1
14	620.5	25.3	333	PCT-US95-08812-4	Sequence 4, Appl1
15	613.5	25.0	333	US-09-214-631-4	Sequence 4, Appl1
16	599.5	24.5	345	US-09-214-631-5	Sequence 5, Appl1
17	489	20.0	89	US-09-214-631-13	Sequence 13, Appl1
18	192	7.8	82	US-09-214-631-11	Sequence 11, Appl1
19	191.5	7.8	82	US-09-214-631-12	Sequence 12, Appl1
20	179	7.3	234	US-08-299-557-5	Sequence 5, Appl1
21	179	7.3	238	US-08-240-194-2	Sequence 2, Appl1
22	179	7.3	238	US-08-453-943-2	Sequence 2, Appl1
23	179	7.3	238	US-09-057-121-2	Sequence 2, Appl1
24	179	7.3	238	US-09-358-734-2	Sequence 2, Appl1
25	176.5	7.2	135	US-08-299-557-7	Sequence 7, Appl1
26	176	7.2	184	US-09-609-324A-2	Sequence 2, Appl1
27	176	7.2	184	US-08-920-440B-2	Sequence 2, Appl1

28	176	7.2	184	4	US-09-173-492-2	Sequence 2, Appl1
29	176	7.2	184	4	US-09-173-133-2	Sequence 2, Appl1
30	176	7.2	184	4	US-09-165-533-2	Sequence 2, Appl1
31	176	7.2	184	5	PCT-US95-12779-2	Sequence 2, Appl1
32	176	7.2	184	5	PCT-US95-15781-2	Sequence 2, Appl1
33	176	7.2	209	1	US-08-455-001-2	Sequence 2, Appl1
34	176	7.2	209	4	US-08-308-814-2	Sequence 2, Appl1
35	176	7.2	209	5	PCT-US95-11669-2	Sequence 2, Appl1
36	175.5	7.2	213	1	US-09-609-324A-10	Sequence 10, Appl1
37	175.5	7.2	213	2	US-08-920-440B-10	Sequence 10, Appl1
38	175.5	7.2	213	3	US-09-173-492-10	Sequence 10, Appl1
39	175.5	7.2	213	4	US-09-173-133-10	Sequence 10, Appl1
40	174.5	7.1	233	4	US-09-214-631-7	Sequence 7, Appl1
41	172	7.0	200	4	US-08-455-001-4	Sequence 4, Appl1
42	172	7.0	200	5	PCT-US95-11669-4	Sequence 4, Appl1
43	169.5	6.9	201	1	US-08-240-124-4	Sequence 4, Appl1
44	169.5	6.9	201	1	US-08-453-943-4	Sequence 4, Appl1
45	169.5	6.9	201	2	US-09-057-121-4	Sequence 4, Appl1
46	169.5	6.9	201	4	US-09-358-734-4	Sequence 4, Appl1
47	169	6.9	201	4	US-09-214-631-8	Sequence 8, Appl1
48	168.5	6.8	209	4	US-09-214-631-6	Sequence 6, Appl1
49	167.5	6.8	228	1	US-08-442-248-4	Sequence 4, Appl1
50	167.5	6.8	228	1	US-08-440-815-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1	US-09-214-631-3	Sequence 3, Application US/09214631	
Patent No. 6413730			
GENERAL INFORMATION:			
APPLICANT: Monsanto, Sacha			
APPLICANT: Monsanto, Geraldine			
APPLICANT: Parnon, Tony			
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR			
FILE REFERENCE: 11757.23USMO			
CURRENT APPLICATION NUMBER: US/09/214,631			
EARLIER FILING DATE: 1997-07-04			
EARLIER APPLICATION NUMBER: 60/021,272			
EARLIER FILING DATE: 1996-07-05			
NUMBER OF SEQ ID NOS: 13			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 3			
LENGTH: 340			
TYPE: PRT			
ORGANISM: Homo sapiens			
US-09-214-631-3			
Query Match	75.0%	Score 1837; DB 4; Length 340;	
Best Local Similarity	99.7%	Pred. No. 2.3e-146;	
Matches 337; Conservative	1; Mismatches	0; Indels	0; Gaps
QY	1	MGPHSGPGVGRVALLLGVLGVGLSPYVWNSANKRFAEGGYLYPQIGRDL 60	
DB	1	MGPHSGPGVGRVALLLGVLGVGLSPYVWNSANKRFAEGGYLYPQIGRDL 60	
QY	61	LCPRAPPGHSSPNYEFKLYIVGAQRCEAPAPVLLTCDPDLFTTFQRY 120	
DB	61	LCPRAPPGHSSPNYEFKLYIVGAQRCEAPAPVLLTCDPDLFTTFQRY 120	
QY	121	SPILMGHFFSHDYIIATSDGTREGLSLOGVCLTRGMKVLTVGQSPRGAVPRK 180	
DB	121	SPILMGHFFSHDYIIATSDGTREGLSLOGVCLTRGMKVLTVGQSPRGAVPRK 180	
QY	181	VSMPEKRDGAHSLPEPKENLPDPTSNATSRKAGGLPPSPMPAYVGAAGLALL 240	
DB	181	VSMPEKRDGAHSLPEPKENLPDPTSNATSRKAGGLPPSPMPAYVGAAGLALL 240	

QY 241 GVAGAGAMCRRRRAPSSRRHPPGSGFRGSGSLGIGGGGGMGPRAEPGELIALRG 300  
 DB 241 GVAAGAGAMCRRRRAPSSRRHPPGSGFRGSGSLGIGGGGGMGPRAEPGELIALRG 300  
 QY 301 GAADPFCPHYKXSGDYGHFVYIVQDGPQSPNNIY 338  
 DB 301 GAADPFCPHYKXSGDYGHFVYIVQDGPQSPNNIY 338

## RESULT 2

US-08-436-044-2  
 ; Sequence 2, Application US/08436044  
 ; Patent No. 5624899  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bennett, Brian D.  
 ; APPLICANT: Matthews, William  
 ; TITLE OF INVENTION: HTK LIGAND  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/436,044  
 ; FILING DATE: 05-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/277722  
 ; FILING DATE: 20-JUL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 902D3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/952-9881  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 336 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-436-044-2

Query Match 25.7%; Score 628.5; DB 1; Length 336;  
 Best Local Similarity 41.8%; Pred. No. 5e-45;  
 Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

QY 14 GALLLVGLVSGLSLEPYVWNSANKRFOAGGVTYVPOIGRDLILCPRAPPPGHS 73  
 DB 17 GLMWLCRTAISRSIVLEPIYWNSSKFLPGQGLVLYPOIGRDLILCPKRV--DSKTV 73  
 QY 74 PNYEFYKLYLVGAQGRCEAPAPNLLTCDRPDLIRFTIKFOEYSPNIMGHEFRSH 133  
 DB 74 GQYEVYKLYVMDKQADRCTIKENTPLNLCARPDOVKFTIKFOEYSPNIMGLEFQKX 133  
 QY 134 DYIIAATSDGTREGLESIGGVCLTRGKVLRYGQ--SPRGAVPRKPVSEMMER-DR 190  
 DB 134 DYIIISTNSGSLGIDNOEGGVCCOTRAKILMKVGDASAGSRNNGPTRRPELEGTN 193  
 QY 191 GAASLEPGKENTLPDPPSNATSRGASGPLPPSPVAVAGAGLALLGVAGAGAMC 250  
 DB 194 GRSSTSPFVKNPQSSITDNGSAGHSNNLIGSEVALFAGIASGCIIFIVITLVVLL 253

QY 251 WRRRAKPSRRHPPGSGFRGSGSLGIGGGGGMGPRAEPGELIALRGGAADPFCPH 310  
 DB 254 KYRRHRKHSFOHTTTLSTLATPKRGANN---NGSEPDVITPLR--TADSVFCPH 306  
 QY 311 YEKVSGDYGHFVYIVQDGPQSPNNIY 338  
 DB 307 YEKVSGDYGHFVYIVQDGPQSPNNIY 334

## RESULT 3

US-08-436-054-2  
 ; Sequence 2, Application US/08436054  
 ; Patent No. 5864020  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bennett, Brian D.  
 ; APPLICANT: Matthews, William  
 ; TITLE OF INVENTION: HTK LIGAND  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/436,054  
 ; FILING DATE: 05-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/277722  
 ; FILING DATE: 20-JUL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 902D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/952-9881  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 336 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-436-054-2

Query Match 25.7%; Score 628.5; DB 2; Length 336;  
 Best Local Similarity 41.8%; Pred. No. 5e-45;  
 Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

QY 14 GALLLVGLVSGLSLEPYVWNSANKRFOAGGVTYVPOIGRDLILCPRAPPPGHS 73  
 DB 17 GLMWLCRTAISRSIVLEPIYWNSSKFLPGQGLVLYPOIGRDLILCPKRV--DSKTV 73  
 QY 74 PNYEFYKLYLVGAQGRCEAPAPNLLTCDRPDLIRFTIKFOEYSPNIMGHEFRSH 133  
 DB 74 GQYEVYKLYVMDKQADRCTIKENTPLNLCARPDOVKFTIKFOEYSPNIMGLEFQKX 133  
 QY 134 DYIIAATSDGTREGLESIGGVCLTRGKVLRYGQ--SPRGAVPRKPVSEMMER-DR 190  
 DB 134 DYIIISTNSGSLGIDNOEGGVCCOTRAKILMKVGDASAGSRNNGPTRRPELEGTN 193  
 QY 191 GAASLEPGKENTLPDPPSNATSRGASGPLPPSPVAVAGAGLALLGVAGAGAMC 250  
 DB 194 GRSSTSPFVKNPQSSITDNGSAGHSNNLIGSEVALFAGIASGCIIFIVITLVVLL 253  
 QY 251 WRRRAKPSRRHPPGSGFRGSGSLGIGGGGGMGPRAEPGELIALRGGAADPFCPH 310

```

Db      254 KRRRRHRKRSFOHTTTLTSLSTATKRGNN-----NGSEPSDVIILPLR---TADSVCFPH 306
Qy      311 YEKVSGDYGHPPYIVIQDGPPOSPPIIY 338
Db      307 YEKVSGDYGHPPYIVIQEMPPOSPPIIY 334

```

RESULT 4  
PCT-US95

```

1      Sequence 2, Application PC/TUS9508812
2      GENERAL INFORMATION:
3      APPLICANT: Genetech, Inc.
4      TITLE OF INVENTION: HTK LIGAND
5      NUMBER OF SEQUENCES: 7
6      CORRESPONDENCE ADDRESS:
7      ADDRESSEE: Genetech, Inc.
8      STREET: 460 Point San Bruno Blvd
9      CITY: South San Francisco
10     STATE: California
11     COUNTRY: USA
12     ZIP: 94080
13     COMPUTER READABLE FORM:
14     MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
15     COMPUTER: IBM PC compatible
16     OPERATING SYSTEM: PC-DOS/MS-DOS
17     SOFTWARE: patin (Genetech)
18     CURRENT APPLICATION DATA:
19     APPLICATION NUMBER: PCT/US95/08812
20     FILING DATE:
21     CLASSIFICATION:
22     PRIOR APPLICATION DATA:
23     APPLICATION NUMBER:
24     FILING DATE:
25     ATTORNEY/AGENT INFORMATION:
26     NAME: Lee, Wendy M.
27     REGISTRATION NUMBER: 00,000
28     REFERENCE/DOCKET NUMBER: 902PCT
29     TELECOMMUNICATION INFORMATION:
30     TELEPHONE: 415/225-1994
31     TELEFAX: 415/952-9881
32     TELEX: 910/371-7168
33     INFORMATION FOR SEQ ID NO: 2:
34     LENGTH: 336 amino acids
35     TYPE: amino acid
36     TOPOLOGY: linear
37     PCT-US95-08812-2

```

Query Match	25.7%;	Score 628.5;	DB 5;	Length 336;
Best Local Similarity	41.8%;	Pred. No. 5e-45;		
Matches 137, Conservative	49;	Mismatches 129;	Indels 13;	Gaps 5

```

Oy 14 GALLLGLVGLVSGLSLSPVYVNSANKRFQKQEGGVLYLPQIGDRDLTLCPAPRPSPHS 73
    14 :::::
Db 17 GLAMVLCTRLASRSIVLEPIYVNSNSKFLPQGGVLYLPQIGDRDLTLCPKVC---DSKTV 73
    17 :::::
Oy 74 PNYEYKYLIVGAGGARGCEAPAPANLLLTDDRPDLRFITIKFOEYSPNMGHEPFSHH 133
    74 :::::
Db 74 GQYEYKYLIVMDKQADACTIKKENTPLINCARPDQVYKFIITKQEBFNMGLGFQKXK 133
    74 :::::
Oy 134 DYIIATIDGTFREGLESIGGYCYLTRGNKVLIRGQ--SPGGVAPRKPVSEMER--DR 199
    134 :::::
Db 134 DYIIITSTNSGLEIGDNOBGVCCQTRANKILIRKQGDSSAGSAPNRGQTRPRPELEGTN 199
    134 :::::
Oy 191 GAHSLSPEKENTLPGDPTSNATSRGAEGLPPSPMPAVAGAAGLIALLLGVAGAAGMC 255
    191 :::::
Db 194 GRSSTSPPVKPNPSSSTIDGNSAGHSNNLLGSEVALPAGIASCGIIFTVITITLVLL 255
    194 :::::
Oy 251 WRRRAKTSSESHPGSPFGROGSLGLOGGGGMPREAPGELGIALRGGAADPFCPH 310
    251 :::::
Db 254 KYRRHRKMSPHHTITLTLSTLATPKRGNN---NSGSESDVITIPN---TASVCPH 306
    254 :::::

```

```

QY      311 YEKVSGDYGHPVYIVODGPPQSPNITY 338
          ||||| | | | | |
          |||||:|||||
          |||||
Db      307 YEKVSGDYGHPVYIVQEMPPQSPNITY 334

```

RESULT 5  
US-08-213-403-2

```

? Sequence 2, Application US/08213403
? Patent No. 5512457
?
? GENERAL INFORMATION:
? APPLICANT: Lyman, Stewart D.
? APPLICANT: Beckmann, M. Patricia
? APPLICANT: Baum, Peter R
? APPLICANT: Carpenter, Melissa
? TITLE OF INVENTION: No. 5512457e1 Cytokine Designated elk Ligand
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Immunex Corporation
? STREET: 51 University Street
? CITY: Seattle
? STATE: Washington
? COUNTRY: USA
? ZIP: 98101
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: Apple Macintosh
? OPERATING SYSTEM: Apple System 7.1
? SOFTWARE: Microsoft Word for Apple, Version 5.1a
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/213,403
? FILING DATE: 15-MAR-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/977,693
? FILING DATE: 13-NOV-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Seese, Kathryn A.
? REGISTRATION NUMBER: 32,172
? REFERENCE/DOCKET NUMBER: 2807-A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206) 587-0430
? TELEFAX: (206) 233-0644
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 346 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
? US-08-213-403-2

```

	Query Match	25.4%	Score 623	DB 1	Length 346
	Best Local Similarity	39.2%	Pred. No. 1.5e-44		
	Matches 143	Conservative 48	Mismatches 116	Indels 58	Gaps 9

Qy	8	PGGRVAGLLLLGLVGVSG-----SLPYPVNSANKRFQAGGVLYVPOIQIDRLDL	61
Db	4	PGGRMLGTVLVAVMVMLCRLATPLAKNLEPVMSSLNPFSLSGKGVIVYPIKIDKXDLII	63
Qy	62	CPRARPPQPHSSPVYEEYKYLYINGAGAGRCCEAPPAVNLITLQDRPDLDIRFTFKPQYS	121
Db	64	CPRAEAGRP-----YETTKYLVAPEDGAAACSYLDVNVVLVTCNRPGEIRFTFKPQFS	118
Qy	122	PNLNGHEFRSHHDYIYIATSDGTRFEGLESLAQGYCLTRGMKVLRLVQSPRGGAVPKPKV	181
Db	119	PNYNGLEKPKKHHDYITSTSTSGSLBGELENGGVCRRRTKILIMKVDQDPNAVTPPEQIT	176
Qy	182	SEMPMERDGRGAHSLP-PGKENTLPDGLTSNATSGKAEGPLRPPEMPRAVAGAGLA----	23
Db	179	SRPSKEDANTVKNATQAPGSRGSLGSDGDKHETVNOEKGSP-----GASGGSSDDPD	23
Qy	237	-----LILLGVAGAGA-----MCRVRRRAPVSSESRRHPGSPFGGSGSLGL	277
Db	232	GFENSKVALFAAVAGGVIFLLIIIFLTVLLLKLRKRRHRTQO-----RAALSL	282

QY 278 ----GGGCGMGPRAEPGEELGIALRGGAADPPCPHYEKVSGDYGHPIVTVODGPPQSP 333  
DB 283 STLASPRGSGGTACTEPESDIIIPLR---TTENNYCPHYEKVSGDYGHPIVTVODGPPQSP 339  
QY 334 PNIYY 338  
DB 340 ANIYY 344

RESULT 6  
US-08-458-077-2  
Sequence 2, Application US/08458077  
Patent No. 5627267  
GENERAL INFORMATION:  
APPLICANT: Lyman, Stewart D.  
APPLICANT: Beckmann, M. Patricia  
APPLICANT: Baum, Peter R.  
APPLICANT: Carpenter, Melissa  
TITLE OF INVENTION: No. 5627267el Cytokine Designated elk Ligand  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,077  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/213,403  
FILING DATE: 15-MAR-1994  
APPLICATION NUMBER: US 07/977,693  
FILING DATE: 13-NOV-1992  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2807-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 346 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-458-077-2

Query Match 25.4%; Score 623; DB 1; Length 346;  
Best Local Similarity 39.2%; Pred. No. 1.5e-44;  
Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGRVVGALLLGLVLSG-----SLRPVYNSANKRQAEGVYVPOIGDRLDL 61  
DB 4 PGGRVVGALLLGLVLSG-----SLRPVYNSANKRQAEGVYVPOIGDRLDL 63

QY 62 CPRARPGPHSSPNYEFYKLYLVGAQGRCEAPPAENLLITCDRPUJDRFTIKQERS 121  
DB 64 CPRARPGPHSSPNYEFYKLYLVGAQGRCEAPPAENLLITCDRPUJDRFTIKQERS 118

QY 122 PNLWGHFRSHDYIATSDGTREGESLQGGVCLTRGKVKLLRVGQSRGGAIVRKPV 181  
DB 119 PNLWGHFRSHDYIATSDGTREGESLQGGVCLTRGKVKLLRVGQSRGGAIVRKPV 178

QY 122 SEMPERDRGAHSLE-PGKENLPGDPTSNATSRGAEGLPPEGMPAVAGAAGLA----- 236  
DB 179 SRPSKEADNTVKNATQAVSGRSLGDSGKHETVQBEKSGP-----GASGGSSGDPD 231

QY 237 -----LILLGVAGAGA-----MCWRRAKPSRPHPGSGFRGSGSL 277  
DB 232 GFENSKVALFAAVGACVIFLLIIIFLVLLKLRKRRKRTQ-----RAALSL 282

QY 278 ----GGGCGMGPRAEPGEELGIALRGGAADPPCPHYEKVSGDYGHPIVTVODGPPQSP 333  
DB 283 STLASPRGSGGTACTEPESDIIIPLR---TTENNYCPHYEKVSGDYGHPIVTVODGPPQSP 339  
QY 334 PNIYY 338  
DB 340 ANIYY 344

RESULT 7  
US-08-460-741-2  
Sequence 2, Application US/08460741  
Patent No. 5670625  
GENERAL INFORMATION:  
APPLICANT: Lyman, Stewart D.  
APPLICANT: Beckmann, M. Patricia  
APPLICANT: Baum, Peter R.  
APPLICANT: Carpenter, Melissa  
TITLE OF INVENTION: No. 5670625el Cytokine Designated elk Ligand  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,741  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/213,403  
FILING DATE: 15-MAR-1994  
APPLICATION NUMBER: US 07/977,693  
FILING DATE: 13-NOV-1992  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2807-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 346 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-460-741-2

Query Match 25.4%; Score 623; DB 1; Length 346;  
Best Local Similarity 39.2%; Pred. No. 1.5e-44;  
Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGRVVGALLLGLVLSG-----SLRPVYNSANKRQAEGVYVPOIGDRLDL 61  
DB 4 PGGRVVGALLLGLVLSG-----SLRPVYNSANKRQAEGVYVPOIGDRLDL 63

QY	62	CPRAAPPEPHSSPNVNEFYKTLVVGGAQGRCEAPAPLTLTQCRPLDPLFTKPEEYS	121
Db	64	CPRAAGRP-----YEVYKTLVLRPEGMAAASCTVLDNPVLVTCNRPGEIRFTIKKPEFS	118
QY	122	ENLMWHERSHHDVYITLITSDPTDEGLSELOGVCLTRGMKVTLLRVGQSPRGCAVPPKRV	181
Db	119	PNYMGLEPKKHHDITLITSTSGSDEGLENRGGVCRTRTKMLIKVQDDPNAVTPPEGLTT	178
QY	182	SEMPWERDGAASHLE-PGKENLPEDPTSNATSRGAEGPLRPSPMAVAGAAGLA----	236
Db	179	SRPSKEADNTVMATQARAGSRGSDSDPKRIETVNOEKSQP-----CASGGSSGDPD	231
QY	237	-----LILLLEVAAGGA-----MCRRRRAKPSBSRHPGSPFGRGSLG	277
Db	232	GFENSKVALFAAVGAGCYTFLITITFLTVLLTKRKRRHRTQO-----RAALSL	282
QY	278	-----GGGGGMPREAREPEGLDIALRGGAADPRPCFHYEKVSGDYGHPRVYIVDGPQSP	333
Db	283	STLASPKSGSGSTAGTEPDIITPLR-----TENNVCFHYEKVSGDYGHPRVYIVDGPMPQSP	339
QY	334	PNITYY 338	
Db	340	ANITYY 344	

US-08-299-567-6

Query Match 25.4%; Score 623; DB 1; Length 346;  
 Best Local Similarity 39.2%; Pred. No. 1.5e-44;  
 Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGRVAGALLLGLVGVSGP-----SLPPVWNSANKRFOAEGGVLYPQIGDLIDL 61  
 Db 4 PGGRVAGALLLGLVGVSGP-----SLPPVWNSANKRFOAEGGVLYPQIGDLIDL 61  
 QY 62 CPRAAPGPHSSPNVEYKYLYVGAQGRCEAPPAENLLTCDRDLRFTIKFOEYSH 121  
 Db 64 CPRAEAGSP-----YEEYKLYLVPEQAAACSTVLDPNVLTGCRPEQETIRFTIKFOES 118  
 QY 122 PNLGHEFRSHHDYIATSDGTRREGLESIOGGVCLTRGKVLRLRVQSPRGAVPRKPV 181  
 Db 119 PNLGHEFRSHHDYIATSDGTRREGLESIOGGVCLTRGKVLRLRVQSPRGAVPRKPV 178  
 QY 182 SEMEMERDGAHSLP-PGKENLPDPTSNATSGAEGPLPPSPMPAVAGAGLA----- 236  
 Db 179 SRPEKADNTYKATQAPGRSGSLGSDGKHETVNOEXSGP-----GASGSSGDDP 231  
 QY 237 -----LILGVAAGAG-----MCRRRRAKPSRHPGSGFRGSLGL 277  
 Db 232 GFENSKYALFAVAGAGCVIFLLIIFLTVLLKLRKHRRKHTQ-----FAAALST 282  
 QY 278 -----GGGGGMPRAEPGELGIALRGGAADPPFCPEYKVSQDPYGHPIYVODGPPSP 333  
 Db 283 STLASPKGSGTAGTEPSDITILPLR---TTENYCPHYEKVSGDYGHPIYVQEMPPOSF 339  
 QY 334 PNYY 338  
 Db 340 ANIYY 344

RESULT 10  
 US-08-436-044-4  
 ; Sequence 4, Application US/08436044  
 ; Patent No. 562489  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bennett, Brian D.  
 ; APPLICANT: Matthews, William  
 ; TITLE OF INVENTION: HTK LIGAND  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/436,044  
 ; FILING DATE: 05-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/277722  
 ; FILING DATE: 20-JUL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 902D3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-1994  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 333 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-436-044-4

Query Match 25.3%; Score 620.5; DB 1; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 2.3e-44;  
 Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

QY 14 GALLLGLVGLVSGISLEPPVWNSANKRFOAEGGVLYPQIGDLIDLCPRAAPGPHSS 73  
 Db 14 GALLLGLVGLVSGISLEPPVWNSANKRFOAEGGVLYPQIGDLIDLCPRAAPGPHSS 73  
 QY 74 PNVEYKYLYVGAQGRCEAPPAENLLTCDRDLRFTIKFOEYSPNLGHEFRSH 133  
 Db 71 GQYETKYVWYDKQADRCTTKENTPLNKAEPDQIKFTIKFOEYSPNLGHEFRSH 130  
 QY 134 DYIITATSDGTRREGLESIOGGVCLTRGKVLRLRVQ--SPRGAVPRKPVSEMER-DR 190  
 Db 131 DYIITATSDGTRREGLESIOGGVCLTRGKVLRLRVQ--SPRGAVPRKPVSEMER-DR 190  
 QY 191 GAHSLPEKXENLPDPTSNATSGAEGPLPPSPMPAVAGAGLALLLGVAGAGAMC 250  
 Db 191 GASTTSPFVKNPSSSTGNSAGHGNNTIGSEVALFAGIASCTITVITLVLL 250  
 QY 251 WRRRAKPSRHPGSGFRGSGSLGCGGGMPPRAEPGELGIALRGGAADPPFCPH 310  
 Db 251 KYRRHRKHSQHTTTLSTLSTLTPKSGNN---NSEPSDITILPLR---TASVCPH 303  
 QY 311 YEKVSGDYGHPIYVODGPPSPNIYY 338  
 Db 304 YEKVSGDYGHPIYVQEMPPOSANIYY 331

RESULT 11  
 US-08-436-054-4  
 ; Sequence 4, Application US/08436054  
 ; Patent No. 5864020  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bennett, Brian D.  
 ; APPLICANT: Matthews, William  
 ; TITLE OF INVENTION: HTK LIGAND  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/436,054  
 ; FILING DATE: 05-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/277722  
 ; FILING DATE: 20-JUL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 902D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-1994  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 333 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
US-08-436-054-4

```

Query Match	25.3%	Score	620.5	DB	2	Length	333
Best Local Similarity	40.5%	Pred. No.	2.3e-44				
Matches	133	Conservative	52	Mismatches	130	Indels	13
						Gaps	5

```

QY 14 GALLMLGVLGVISGSLSEPFYVWVNSANKRFQAEQGVLYLPQJGDLGLDCEAPRPPGSHS 73
Db 14 GVLAMLCRTALSKSIVLEPIYWNSSNSKFLPGQGVLYLPQJGDLMDIICPV---DSKV 70
QY 74 PNEYVKLYLVGAGQGRCEAPRANMLLTCDRDLRFLPTIKQVSPNIMGHEPFSHH 133
Db 71 GQYEYIKYVMWDKQADCCITKKENTPLNCAKPDQDITKFIKKQESSPNIMGHEPQNK 130
QY 134 DYYIATSDGTRREGLSHSQGVCLTRGMKVLRLRQO--SPRGAVPRKPSSEMEMER-DR 150
Db 131 DYYIISTNGSLGLELDNGEGGVCCQRAKMLIMKVGDQASASGSTRNDDPTRRPELKGTN 150
QY 191 GAASHLEEGKENLBDPTFSNATSKGABEPLRPBSMPAVAGAGCTALLILGVAGAGAMC 250
Db 191 GRSSSTSEFVFPNPPSSSTDGNSAGHSNGNLLIGSEVALPAGIASCIIIFYIITLVLL 250
QY 251 WRRRRAKPESERHPQPGSFCRGSGLGCGGGMGPRAEPGELIALRGGAADPFCPH 310
Db 251 KYRRRHGRHSPOHTTTTSLTAPKXSGNN---NSEPEDIITPLR---TADSVCPH 303
QY 311 YEKVSGDYGHPIYLVQDGPPOSPPIITY 338
Db 304 YEKVSGDYGHPIYLVQEMPPOSPPIITY 331

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	25.3%;	Score 620.5;	DB 4;	Length 333;
Best Local Similarity	40.5%;	Pred. No. 2.3e-44;		
Matches 133;	Conservative 52;	Mismatches 130;	Indels 13;	Gaps 5;

[illegible]

```

1      RESULT 13
2      PCT-US95-08534-2
3      Sequence 2, Application PC/TUS9508534
4      GENERAL INFORMATION:
5      APPLICANT: Immunex Corporation
6      TITLE OF INVENTION: Novel Cytokine Desig
7      NUMBER OF SEQUENCES: 3
8      CORRESPONDENCE ADDRESS:
9      ADDRESSEE: Immunex Corporation
10     STREET: 51 University Street
11     CITY: Seattle
12     STATE: Washington
13     COUNTRY: US
14     ZIP: 98101
15     COMPUTER READABLE FORM:
16     MEDIUM TYPE: Floppy disk
17     COMPUTER: Apple Macintosh
18     OPERATING SYSTEM: Apple 7.1
19     SOFTWARE: Microsoft Word, Version 5.1a
20     CURRENT APPLICATION DATA:
21     APPLICATION NUMBER: PCT/US95/08534
22     FILING DATE: 06-JUL-1995
23     CLASSIFICATION:
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER: US 08/271,948
26     FILING DATE: 08-JUL-1994
27     CLASSIFICATION:
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Anderson, Kathryn A.
30     REGISTRATION NUMBER: 32,172
31     REFERENCE/DOCKET NUMBER: 2823-WO
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (206) 587-0430
34     TELEFAX: (206) 233-0644
35     TELEX: 756822
36     INFORMATION FOR SEQ ID NO: 2:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 333 amino acids
39     TYPE: amino acid
40     TOPOLOGY: linear
41     MOLECULE TYPE: protein
42     CDT-US95-08534-2

```



[illegible]

```

RESULT 15
US-09-214-631-4
; Sequence 4, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-4

Query Match      25.0%; Score 613.5; DB 4; Length 333;
Best Local Similarity 40.5%; Pred. No. 9e-44;
Matches 133; Conservative 54; Mismatches 128; Indels 13; Gaps 6;

QY      14  GALLILGVLGIVSGSLSEPPYWNSSANKRFQAEGGYVLYPOIGDRDLDTCPAARDPPGHSS 73
DB      14  GVLNMLCTTAISKSTIVLEPIYWNSSNKKFLPGQGLVLPQGDGLDITCPKY--DSKTV 70
QY      74  PNYEYFKLYIVGAGARRCEAPPAENLLLTCDRDLDTFTIKFOEYSPNIMGHEFRSH 133
DB      71  GQYEYKYKVMYMDKQADRCITIKENTPLTNCAPKDPDQIKFTIKFOEYSPNIMGHEFRSH 130
QY      134  DYIIATSDGTREGHLSQGGVCLTRGKLYLRYGO--SPRGAVPRKVSMP--MER-DR 190
DB      131  DYIIISTNSGSLGIDNQGVCQTRANKILMKVGDQSSAGS--PNKDDPTIRPELEAGTN 190
QY      191  GAASHLEPGKENTLPDGPSPNATSRGAEGLPPPEMPAVAGAGATALLLGVAGAGAMC 250
DB      191  GRSSTTFPVKPNPGSSIDTNGSAGHSGNNILGSEVALFAGIASGCIIFVITITLVYLL 250
QY      251  WRRRAKSESRHQPQSGFGGSGSLGAGGGGGMPPRAEPPELGTALRGGAADPPPCPH 310

```

```
Db 251 KYRRHRKPSPHHTTTLSTLATPKRSNN-----NGSEPSDIIPLR---TADSVCFPH 303
QY 311 YKVSQDYGHPYIVODGPPSPNNITY 338
Db 304 YKVSQDYGHPYIVODGPPSPNNITY 331
```

```
RESULT 16
US-09-214-631-5
; Sequence 5, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawsan, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-5
```

```
Query Match 24.5%; Score 599.5; DB 4; Length 345;
Best Local Similarity 38.4%; Pred. No. 1.4e-42;
Matches 140; Conservative 50; Mismatches 116; Indels 59; Gaps 10;
```

```
QY 8 PGQVRV-----ALLLGVTVSGV--SLEPYTNSANKRFOAEGVLYTQIGPRDL 61
Db 4 PGQRMLGKVLVAMVWALCRLATPLAKLEPVSWSLNPKFLSGGLVYIKIGKLDII 63
QY 62 CPRAAPPGHSPNTEFYLYLVGAQGRCEAPAPNLLITCDRPPDLFTTIFQEYS 121
Db 64 CRRPAEGR-----YEVYLYVIREPQAAAGCTVADPMVLTQCRPBEIIFTIKQFSS 118
QY 122 PVLKHEPSSHDIYIATSDGTREGLESLOGVCLTRGMKVLTRVGSPPGCAVPRPV 181
Db 119 PVMGLIEFKHDHYITSTNSNGSLGLENRGGVCRTMTKIMKVGDPMNAVTPPEQLTT 178
QY 182 SEMPERDRGAHSLR-PCKENLPDDPTSMATSRGAEPLPPSPMAYAGAAGLA---- 236
Db 179 SRPSKEADNTVMATQAPPSRGLSDSGKHEFTVQEKSGP-----GASGSSGDPD 231
QY 237 -----LLLLGYAAGGA-----MCMRRRAKPSERHPPGSGFRGGSIGL 277
Db 232 GFENSKVALFAVAGGCVFTIITITLVTLKLPKRHKHTQ-----RAALSL 281
QY 278 -----GGGGMGPREAPGELGIALRGGAADPPFCPIYEKISGDYGHVYIVODGPPSP 333
Db 282 STIASPKGSGTAGTEPSDIIPL--FTTENNYCPHYEKVSGDYGHVYIVODEMPQSP 338
QY 334 PNITY 338
Db 339 ANITY 343
```

```
RESULT 17
US-09-214-631-13
; Sequence 13, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
```

```
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawsan, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-13
```

```
Query Match 20.0%; Score 489; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 4.5e-34;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 252 RRRRAKPSERHPPSGFRGSLTGCGGGMGPRAEPGEIGIALRGGAADPPCPHY 311
Db 1 RRRRAKPSERHPPSGFRGSLTGCGGGMGPRAEPGEIGIALRGGAADPPCPHY 60
QY 312 EKVSQDYGHPYIVODGPPSPNNITY 338
Db 61 EKVSQDYGHPYIVODGPPSPNNITY 87
```

```
RESULT 18
US-09-214-631-11
; Sequence 11, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-11
```

```
Query Match 7.8%; Score 192; DB 4; Length 82;
Best Local Similarity 46.2%; Pred. No. 3.2e-09;
Matches 42; Conservative 9; Mismatches 24; Indels 16; Gaps 3;
```

```
QY 252 RRRRAKPSERHPPSGFRGSLTG-----GGGGMGPREAPGELGIALRGGAADPPF 307
Db 2 RRRRAKHTQ-----RAALSLSTIASPKGSGTAGTEPSDIIPLR---TTENNY 49
QY 308 CPHYEKVSGDYGHVYIVODGPPSPNNITY 338
Db 50 CPHYEKVSGDYGHVYIVODEMPQSPNNITY 80
```

```
RESULT 19
```

```

US-09-214-631-12
; Sequence 12, Application US/09214631
; Patent No. 6413730
;
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-12

```



MOLECULE TYPE: protein  
US-09-057-121-2

Query Match 7.3%; Score 179; DB 2; Length 238;  
Best Local Similarity 28.4%; Pred. No. 1.6e-07;  
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVVGAALLLGLVGLSLPEVYVNSANKRFOAGGYLYVPOIGRDLDCP--R 64  
DB 24 GPG-----GALG-----NRHAYVWNSNQHLLRE-GYTVQVNVNDYLDIYCPHYN 67  
QY 65 ARPGPHSSP-----NVEFYKLYVGAQGRCEAPAPNLLTCDPDL---DLRFTIKF 117  
DB 68 SSGVGPAGGPGPGGAGQVLYVNSNGYRTCNASQGR-RWECNRPHAPHSPIKSEKF 126  
QY 118 QEYSPNLWGHFRSHHDYIATSDGTREGLESLOQGVCLTRGKVLRLVQSPRGAVP 177  
DB 127 QRYSAFSLGYEFHAGHEYIYSTPTNLH-----WKCLR--MKVFVCCASTSHSG--- 174  
QY 178 RKPVSMP-----MERDGAHSLF-----PKENLP 204  
DB 175 EKPVPFTLPQFTMGPNVXINVLEDFEGENPQVPKLEKSIKSTSPKREHLP 223

## RESULT 24

US-09-358-734-2  
Sequence 2, Application US/09358734  
Patent No. 627417  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: Apple Macintosh  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/358,734  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-358-734-2

Query Match 7.3%; Score 179; DB 4; Length 238;  
Best Local Similarity 28.4%; Pred. No. 1.6e-07;  
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVVGAALLLGLVGLSLPEVYVNSANKRFOAGGYLYVPOIGRDLDCP--R 64  
DB 24 GPG-----GALG-----NRHAYVWNSNQHLLRE-GYTVQVNVNDYLDIYCPHYN 67  
QY 65 ARPGPHSSP-----NVEFYKLYVGAQGRCEAPAPNLLTCDPDL---DLRFTIKF 117  
DB 68 SSGVGPAGGPGPGGAGQVLYVNSNGYRTCNASQGR-RWECNRPHAPHSPIKSEKF 126  
QY 118 QEYSPNLWGHFRSHHDYIATSDGTREGLESLOQGVCLTRGKVLRLVQSPRGAVP 177  
DB 127 QRYSAFSLGYEFHAGHEYIYSTPTNLH-----WKCLR--MKVFVCCASTSHSG--- 174  
QY 178 RKPVSMP-----MERDGAHSLF-----PKENLP 204  
DB 175 EKPVPFTLPQFTMGPNVXINVLEDFEGENPQVPKLEKSIKSTSPKREHLP 223

## RESULT 25

US-08-299-567-7  
Sequence 7, Application US/08299567  
Patent No. 5747033  
GENERAL INFORMATION:  
APPLICANT: Davis, et al.  
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL  
TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-6707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/299,567  
FILING DATE: 01-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempster, Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ. ID NO. 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-299-567-7

Query Match 7.2%; Score 176.5; DB 1; Length 135;  
Best Local Similarity 31.9%; Pred. No. 1.2e-07;  
Matches 43; Conservative 14; Mismatches 73; Indels 5; Gaps 2;

QY 33 VYWSANKRFOAGGYLYVPOIGRDLDCP--RAPPGHSSBNTEFYKLYVGAQGR 90  
DB 1 VYWSNSNPKFYRXEGYITIXVXXNDYLDIICPHYEXXXXXXXAGXECYXLYVXXEXXX 60  
QY 91 RGAAPAPNLLTCDP---DLDRFTIKQEYSPNLWGHFRSHHDYIATSDGTREG 147

```

; OPERATING SYSTEM: System 7.6
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-173-492-2

Query Match 7.2%; Score 176; DB 4; Length 184;  
Best Local Similarity 29.3%; Pred. No. 2e-07;  
Matches 59; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYNSANKRFQAE-----GGVLYPQIGDRIDLCPARPFGPHSSPNVEFYKLYLVGA 87  
DB 10 VYNSNSNRFQVSAVGDGGYTVESINDYLDICPHYGAPLP-PAERMERYYILYMWNGE 68  
QY 88 QGRCEAPAPNLLLTCDRPLD---DLRFTIKPOEYSPNLMGHEFRSHHDYIIATSDGT 144  
DB 69 GHASCDHRQGRFKMECKRPAAPGGLKFSKFLTFPSLGFEPFGHEYYIISATP-- 126  
QY 145 REGLESLOGVCLTRGMKVLIRVQSPRGAVPRKPVSEMPMERDRGAHSLPECKENLP 204  
DB 127 ---PNLVDRCLR--LKYYR-----FTMETLY 149  
QY 205 GDP-----TSNATSRGAG 218  
DB 150 EAPPIFTSNSSCSGLGG 167

## RESULT 29

US-09-173-133-2  
Sequence 2, Application US/09173133  
Patent No. 6232447  
GENERAL INFORMATION:  
APPLICANT: Cerretti, Douglas P.  
TITLE OF INVENTION: Cytokine Designated LERK-6  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Power Macintosh  
OPERATING SYSTEM: System 7.6  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173,133  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,440  
FILING DATE: 29-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Henry, Janis C.  
REGISTRATION NUMBER: 34,347  
REFERENCE/DOCKET NUMBER: 2826-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-173-133-2

Query Match 7.2%; Score 176; DB 4; Length 184;

Best Local Similarity 29.3%; Pred. No. 2e-07;  
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYNSANKRFQAE-----GGVLYPQIGDRIDLCPARPFGPHSSPNVEFYKLYLVGA 87  
DB 10 VYNSNSNRFQVSAVGDGGYTVESINDYLDICPHYGAPLP-PAERMERYYILYMWNGE 68  
QY 88 QGRCEAPAPNLLLTCDRPLD---DLRFTIKPOEYSPNLMGHEFRSHHDYIIATSDGT 144  
DB 69 GHASCDHRQGRFKMECKRPAAPGGLKFSKFLTFPSLGFEPFGHEYYIISATP-- 126  
QY 145 REGLESLOGVCLTRGMKVLIRVQSPRGAVPRKPVSEMPMERDRGAHSLPECKENLP 204  
DB 127 ---PNLVDRCLR--LKYYR-----FTMETLY 149  
QY 205 GDP-----TSNATSRGAG 218  
DB 150 EAPPIFTSNSSCSGLGG 167

## RESULT 30

US-09-165-533-2  
Sequence 2, Application US/09165533  
Patent No. 6268482  
GENERAL INFORMATION:  
APPLICANT: Cerretti, Douglas P.  
TITLE OF INVENTION: Cytokine Designated as LERK-6  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: System 7.1  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/165,533  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/538,709  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Malaska, Stephen L.  
REGISTRATION NUMBER: 33,655  
REFERENCE/DOCKET NUMBER: 2826  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-165-533-2

Query Match 7.2%; Score 176; DB 4; Length 184;  
Best Local Similarity 29.3%; Pred. No. 2e-07;  
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYNSANKRFQAE-----GGVLYPQIGDRIDLCPARPFGPHSSPNVEFYKLYLVGA 87  
DB 10 VYNSNSNRFQVSAVGDGGYTVESINDYLDICPHYGAPLP-PAERMERYYILYMWNGE 68  
QY 88 QGRCEAPAPNLLLTCDRPLD---DLRFTIKPOEYSPNLMGHEFRSHHDYIIATSDGT 144  
DB 69 GHASCDHRQGRFKMECKRPAAPGGLKFSKFLTFPSLGFEPFGHEYYIISATP-- 126

QY 145 REGLESLQGVCLTGMKVLTVGSPRGAVPRKPVSEMPMERDGAHSLPECKENLP 204  
Db 127 ----PNTVDRPCLR--LKYVYR-----PNTETLY 149  
QY 205 GDP---TSNATSRGAEG 218  
Db 150 EAPERIFTSNNSCSGLGG 167

## RESULT 31

PCT-US95-12779-2  
; Sequence 2, Application PC/TUS9512779  
; GENERAL INFORMATION:  
; APPLICANT: Cerretti, Douglas P.  
; TITLE OF INVENTION: Cytokine Designated LERK-6  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: System 7.1  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/12779  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Malaska, Stephen L.  
; REGISTRATION NUMBER: 32,655  
; REFERENCE/DOCKET NUMBER: 2826-WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-12779-2

Query Match 7.2%; Score 176; DB 5; Length 184;

Best Local Similarity 29.3%; Pred. No. 2e-07; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYWNANKRFOAE-----GGVLYVQIGDRDLCPRARPPGPHSSPNVEFYKLYLVGA 87  
Db 10 VYWNNSNRFQYSAVDDGGGYTVESINDYLDYCPHGAFLP-PAERERLYILYMNGE 68  
QY 88 QGRCEAPPAENLLITCDRPDI--DLFTTFQSYSPNLMGHEFRSHHDYIITSDGT 144  
Db 69 GHASCDHRQGRKWEKCNRPAPAGPGLKFSKFOLETFEFLGFEPFRPGHEYYIATP-- 126  
QY 145 REGLESLQGVCLTGMKVLTVGSPRGAVPRKPVSEMPMERDGAHSLPECKENLP 204  
Db 127 ----PNTVDRPCLR--LKYVYR-----PNTETLY 149

QY 205 GDP---TSNATSRGAEG 218  
Db 150 EAPERIFTSNNSCSGLGG 167

## RESULT 32

PCT-US95-15781-2  
; Sequence 2, Application PC/TUS9515781  
; GENERAL INFORMATION:

; APPLICANT: Cerretti, Douglas P.  
; TITLE OF INVENTION: Cytokine Designated Lerk-7  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: System 7.1

; SOFTWARE: Patent Release #1.0, Version #1.25

; APPLICATION NUMBER: PCT/US95/15781

; FILING DATE: 05-DEC-1995

; CLASSIFICATION:

; APPLICATION NUMBER: US 08/351,025

; FILING DATE: 06-DEC-1994

; CLASSIFICATION:

; APPLICATION NUMBER: US 08/396,946

; FILING DATE: 01-MAR-1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Anderson, Kathryn A.

; REGISTRATION NUMBER: 32,172

; REFERENCE/DOCKET NUMBER: 2829-WO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 587-0430

; TELEFAX: (206) 233-0644

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 184 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US95-15781-2

Query Match 7.2%; Score 176; DB 5; Length 184;

Best Local Similarity 29.3%; Pred. No. 2e-07; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYWNANKRFOAE-----GGVLYVQIGDRDLCPRARPPGPHSSPNVEFYKLYLVGA 87  
Db 10 VYWNNSNRFQYSAVDDGGGYTVESINDYLDYCPHGAFLP-PAERERLYILYMNGE 68  
QY 88 QGRCEAPPAENLLITCDRPDI--DLFTTFQSYSPNLMGHEFRSHHDYIITSDGT 144  
Db 69 GHASCDHRQGRKWEKCNRPAPAGPGLKFSKFOLETFEFLGFEPFRPGHEYYIATP-- 126  
QY 145 REGLESLQGVCLTGMKVLTVGSPRGAVPRKPVSEMPMERDGAHSLPECKENLP 204  
Db 127 ----PNTVDRPCLR--LKYVYR-----PNTETLY 149

QY 205 GDP---TSNATSRGAEG 218  
Db 150 EAPERIFTSNNSCSGLGG 167

US-08-455-001-2

; Sequence 2, Application US/08455001

; Patent No. 5795734

; GENERAL INFORMATION:

; APPLICANT: Planagan, John G.

; TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:



```

ADDRESSSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,001
FILING DATE: 31 MAY 1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 56,709
REFERENCE/DOCKET NUMBER: HMT-011CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-455-001-2

Query Match          7.2%; Score 176; DB 1; Length 209;
Best Local Similarity 29.3%; Pred.No. 2.4e-07;
Matches 58; Conservative 19; Mismatches 69; Gaps 52; Gaps 7

QY      33 VYNNANKRRQAE-----GGVYLPIQIGRDLDCPRAPGPHSSPNVEFYKLYLGYGA 87
       |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      35 VYNNRSRPQRQVAVGDGGGYTWSINDYDITCPHYGALP-PALMERIYIIYNNGE 93
QY      88 QGRRCRAPPAENLLTCDRPDL--DLRFITFGYSNLMGHFRSHMDYIIATSDGT 144
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db      94 GHASCDHRQGFRKWECONRPPAPRGFLTKSEKFCLFTFSLGFEFFRGHEYYYSATP-- 151
QY     145 REGLESQQGVCTLRGQKVLRYGSPRGAGVPRAFPVSEMNMENDRGASLEPGKENLP 204
       .:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db     152 ----PNLVDPCLR-LKYIVR-----PYNETLY 174
QY     205 GDP----TSNATSRGAG 218
Db     175 EAPPEPTSNSSCSGLGG 192

RESULT 34
US-08-308-814-2
Sequence 2, Application US/08308814
Patent No. 6268476
GENERAL INFORMATION:
APPLICANT: Flanagan, John G.
TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related
TITLE OF INVENTION: There to
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/308,814
: FILING DATE: 19-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709
: REFERENCE/DOCKET NUMBER: HMI-011
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 209 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-308-814-2

Query Match 7.2%; Score 176; DB 4; Length 209;
Best Local Similarity 29.3%; Pred. No. 2,4e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

Cy 33 VYVNSANRFPQAE-----GGYVLIPQLQEDRLDLCPRARPPGPHSSPYVEYKYLIVGCA 87
Db 35 VYVNSNRPFPVSAVGDGGYTVSEVINDYLDICPHYGAPLP-PAERMESYILYMWNGE 93
Cy 88 QGRRCAPAPNPLLTCDRPLD---DLRFITKFEVSPNLMGHEPRSHDYIATSDGT 144
Db 94 GHASDHQGRGFKMECKRPAAPGGPLKFSKPLFTFSLGFETFRPHETYYISATP-- 151
Cy 145 REGESLGGVCLTRGMKVLIRVGQSPRGAVPRKPVSEMPERDRGAHSLPEKENLP 204
Db 152 ---PNLVDRPCLRL-LKVVYR-----PTNETLY 174
Cy 205 GDP----TSNATSRGAE 218
Db 175 EAPETFTSNSSCSGLGG 192

RESULT 35
PCT-US95-11869-2
: Sequence 2, Application PC/TUS9511869
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: EPH Receptor ligands, and Uses Related
: TITLE OF INVENTION: There to
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11869
: FILING DATE: 19-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709
: REFERENCE/DOCKET NUMBER: HMI-011CPCP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 209 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-11869-2

Query Match 7.2%; Score 176; DB 5; Length 209;
Best Local Similarity 29.3%; Pred. No. 2,4e-07;

```





NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HMI-011CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-455-001-4

Query Match 7.0%; Score 172; DB 1; Length 200;  
Best Local Similarity 35.3%; Pred. No. 4.9e-07;  
Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;

QY 15 ALLILGVLGSLGP-----VYNSANKRFOAGGYLYPQIGRLDLCPPRA 65  
DB 7 AALLAATVG-VCVMSDDGKVISDRYAVYNNRSNPRFH-RGDYVEVSIINDYLDIYCPHY 64  
QY 66 RPPGHSPPNYEFYLYLVGAGRCRCEAPPAPNLLTCDRPLD---DLRFTTKFOQYSP 122  
DB 65 EEPPLP--AEKMERVYLVMNVEGHASCDHRKGRKWEKCNRPDSGPIKFSEKQLFTF 122  
QY 123 NMGHERSHDYIATIS 141  
DB 123 FSLGFEFRPGHSHYIISAS 141

RESULT 42  
PCT-US95-11869-4  
Sequence 4, Application PC/TUS9511869  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related  
TITLE OF INVENTION: Thiereto  
NUMBER OF SEQUENCES: 5  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE:  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/11869  
FILING DATE: 19-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HMI-011CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-11869-4

Query Match 7.0%; Score 172; DB 5; Length 200;  
Best Local Similarity 35.3%; Pred. No. 4.9e-07;  
Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;

QY 15 ALLILGVLGSLGP-----VYNSANKRFOAGGYLYPQIGRLDLCPPRA 65  
DB 7 AALLAATVG-VCVMSDDGKVISDRYAVYNNRSNPRFH-RGDYVEVSIINDYLDIYCPHY 64  
QY 66 RPPGHSPPNYEFYLYLVGAGRCRCEAPPAPNLLTCDRPLD---DLRFTTKFOQYSP 122  
DB 65 EEPPLP--AEKMERVYLVMNVEGHASCDHRKGRKWEKCNRPDSGPIKFSEKQLFTF 122

QY 123 NMGHERSHDYIATIS 141  
DB 123 FSLGFEFRPGHSHYIISAS 141

RESULT 43  
US-08-240-124-4  
Sequence 4, Application US/08240124  
Patent No. 5516658  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEX  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,124  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEER, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-240-124-4

Query Match 6.9%; Score 169.5; DB 1; Length 201;  
Best Local Similarity 29.9%; Pred. No. 8e-07;  
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGSLT-EPYVNSANKRFOAGGYLYPQIGRLDLCPPRPPGHSPPNYEFYLY 82  
DB 20 LRGSLSLHVYVYNNSSNRL-LRGDAVVELGANDYLDIVCHYSGPGRPPRP--ETFPALY 76  
QY 83 LVGAGRCRCEAP-PAPNLLTCDRPLDLEFTTKFOQYSPNMGHERSHDYIATIS 141  
DB 77 WYDWPGVSCQAGRPAYKRWVCSLPFGHVQFSKIORFTFSGFEFLPGSETYYIISVP 136  
QY 142 DSTRGLSLGQGVLTGMKVLARVGSPPRGAVPKRPVSEMMEDRGAHSLPEPKE 201  
DB 137 --TPE-----SSGOCL-----RLQVSVCKEKRSBSAHV----- 164

CY 202 NLPDPTSNATS--RGAEGLPPSPMPAVAGAGLALLL 240  
Db 165 ---GSPGESGTSQWRGDPSP-----LCALLL 189

## RESULT 44

US-08-453-943-4  
; Sequence 4, Application US/08453943  
; Patent No. 5738844  
; GENERAL INFORMATION:  
; APPLICANT: BECKMANN, M. P.  
; APPLICANT: CERRETTI, DOUGLAS P.  
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMUNEX CORPORATION  
; STREET: 51 UNIVERSITY STREET  
; CITY: SEATTLE  
; STATE: WASHINGTON  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple System 7.1  
; SOFTWARE: Microsoft Word for Apple, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,943  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/240,124  
; FILING DATE: 09-MAY-1994  
; APPLICATION NUMBER: US 08/161,132  
; FILING DATE: 03-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/114,426  
; FILING DATE: 30-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/109,745  
; FILING DATE: 20-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SEESE, KATHRYN A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2814-C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-453-943-4

Query Match 6.9%; Score 169.5; DB 1; Length 201;  
Best Local Similarity 29.9%; Pred. No. 8e-07;  
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;  
CY 24 LVGSLT-EPYVNSANKRFOAEGGYLYPOIGRLDLCPRAPPGPHSSPNYEFYKLY 82  
Db 20 LRGSSSLRHVVYVNSNPRL-LRGDAVVELGLNDYLDIVCPHYEGPBPBPSP--ETFALY 76  
CY 83 LVGAQGRRCAP-PAPVLLTCDRPLDLRFTIKFQYSPNLMGHERSHNDYIAT 141  
Db 77 WVDWPGVSCAEGPRAYKRVVCSLPFGHVQFSKIQRTFTFSIGFEPLDGETIYIISVP 136  
CY 142 DGTREGLESLOGVCLTRGMKVLRLVGQSPRGAVPRKPYSEMPMERDRGAASLEBQKE 201  
Db 137 --TPE-----SSGOCL-----RLQVSVCKEKRSASAPV----- 164

CY 202 NLPDPTSNATS--RGAEGLPPSPMPAVAGAGLALLL 240  
Db 165 ---GSPGESGTSQWRGDPSP-----LCALLL 189

## RESULT 45

US-09-057-121-4  
; Sequence 4, Application US/09057121  
; Patent No. 5969110  
; GENERAL INFORMATION:  
; APPLICANT: BECKMANN, M. P.  
; APPLICANT: CERRETTI, DOUGLAS P.  
; TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMUNEX CORPORATION  
; STREET: 51 UNIVERSITY STREET  
; CITY: SEATTLE  
; STATE: WASHINGTON  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple System 7.1  
; SOFTWARE: Microsoft Word for Apple, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/057,121  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/240,124  
; FILING DATE:  
; APPLICATION NUMBER: US 08/161,132  
; FILING DATE: 03-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/114,426  
; FILING DATE: 30-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/109,745  
; FILING DATE: 20-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SEESE, KATHRYN A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2814-C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-057-121-4

Query Match 6.9%; Score 169.5; DB 2; Length 201;  
Best Local Similarity 29.9%; Pred. No. 8e-07;  
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;  
CY 24 LVGSLT-EPYVNSANKRFOAEGGYLYPOIGRLDLCPRAPPGPHSSPNYEFYKLY 82  
Db 20 LRGSSSLRHVVYVNSNPRL-LRGDAVVELGLNDYLDIVCPHYEGPBPBPSP--ETFALY 76  
CY 83 LVGAQGRRCAP-PAPVLLTCDRPLDLRFTIKFQYSPNLMGHERSHNDYIAT 141  
Db 77 WVDWPGVSCAEGPRAYKRVVCSLPFGHVQFSKIQRTFTFSIGFEPLDGETIYIISVP 136  
CY 142 DGTREGLESLOGVCLTRGMKVLRLVGQSPRGAVPRKPYSEMPMERDRGAASLEBQKE 201  
Db 137 --TPE-----SSGOCL-----RLQVSVCKEKRSASAPV----- 164

Db 137 --TPE-----SSGQCL-----RLQVSVCKEKRRSSAHV----- 164

Qy 202 NLPDPTSNATS--RGAEGLPPSPMPAVAGAAGLALLL 240

Db 165 ---GSPGSGTSGMRGDTTSP-----LCULLL 189

RESULT 46

US-09-358-734-4

Sequence 4, Application US/09358734

Patent No. 6274117

GENERAL INFORMATION:

APPLICANT: BECKMAN, M. P.

APPLICANT: CERRETTI, DOUGLAS P.

TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE

TITLE OF INVENTION: RECEPTOR HER

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMUNEX CORPORATION

STREET: 51 UNIVERSITY STREET

CITY: SEATTLE

STATE: WASHINGTON

COUNTRY: USA

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Apple System 7.1

SOFTWARE: Microsoft Word for Apple, Version 5.1a

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/358, 734

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/240,124

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/114,426

FILING DATE: 30-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/109,745

FILING DATE: 20-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: SEESE, KATHRYN A.

REGISTRATION NUMBER: 32,172

REFERENCE/DOCKET NUMBER: 2814-C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644

TELEX: 756822

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 201 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-358-734-4

Query Match 6.9%; Score 169.5; DB 4; Length 201;

Best Local Similarity 29.8%; Pred. No. 8e-07;

Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

Qy 24 LVSGISL-EPYVWNSANRFOAEGGYLYPOIGRDLICPRAPPQGHSSPNTEFYKLY 82

Db 20 LRGSLSLHVYVWNSNRL-LRGAVALGLINDYLDIVCPHYGPGPPEGP--ETPALY 76

Qy 83 LVGAGQGRCEAP-PAPVLLITCDRPDLAFTTKQOYSPNLMGHRFSRSHDYIATS 141

Db 77 MDMWGYISCOABGRPRAYKRWVCSLPGHVQFSEKIQFTFSLGFEFLPGEITYIISVP 136

Qy 142 DGTREGLSLGGVCLTRGMKYLRLVQSPRGAVPRKPYSEMMEBERDGAHSLPEGKE 201

Db 137 --TPE-----SSGQCL-----RLQVSVCKEKRRSSAHV----- 164

Qy 202 NLPDPTSNATS--RGAEGLPPSPMPAVAGAAGLALLL 240

Db 165 ---GSPGSGTSGMRGDTTSP-----LCULLL 189

RESULT 47

US-09-214-631-8

Sequence 8, Application US/09214631

Patent No. 6413730

GENERAL INFORMATION:

APPLICANT: HOLLAND, Sacha

APPLICANT: Mbamalu, Geraldine

APPLICANT: Pawson, Tony

TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

FILE REFERENCE: 11757.23USWO

CURRENT APPLICATION NUMBER: US/09/214,631

CURRENT FILING DATE: 1999-03-12

EARLIER APPLICATION NUMBER: PCT/CA97/00473

EARLIER FILING DATE: 1997-07-04

EARLIER APPLICATION NUMBER: 60/021,272

EARLIER FILING DATE: 1996-07-05

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 201

TYPE: PRT

ORGANISM: Homo sapiens

US-09-214-631-8

Query Match 6.9%; Score 169; DB 4; Length 201;

Best Local Similarity 29.8%; Pred. No. 8e-07;

Matches 70; Conservative 19; Mismatches 84; Indels 62; Gaps 11;

Qy 17 LILGVL-----GLVSGSL-EPYVWNSANRFOAEGGYLYPOIGRDLICPRAPP 68

Db 6 LITVMAAFSGPLRGSSSLHVVYVWNSNRL-SLRGAVALGLINDYLDIVCPHYGEP 64

Qy 69 GPHSSPNTEFYKLYLVGAGQGRCEAP-PAPVLLITCDRPDLAFTTKQOYSPNLMG 127

Db 65 GPPEGP--ETPALYVMDWGYISCOABGRPRAYKRWVCSLPGHVQFSEKIQFTFSLG 122

Qy 128 EFRSHHDYIITSDGTREGSLGGVCLTRGMKYLRLVQSPRGAVPRKPYSEMME 187

Db 123 EFLPGETIYIISVP--TPE-----SSGQCL-----RLQVSVCKE 155

Qy 188 RDRGAHSLPEKENVLPDPTSNATS--RGAEGLPPSPMPAVAGAAGLALLL 240

Db 156 RKSESAPV-----GSPGSGTSGMRGDTTSP-----LCULLL 189

RESULT 48

US-09-214-631-6

Sequence 6, Application US/09214631

Patent No. 6413730

GENERAL INFORMATION:

APPLICANT: HOLLAND, Sacha

APPLICANT: Mbamalu, Geraldine

APPLICANT: Pawson, Tony

TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

FILE REFERENCE: 11757.23USWO

CURRENT APPLICATION NUMBER: US/09/214,631

CURRENT FILING DATE: 1999-03-12

EARLIER APPLICATION NUMBER: PCT/CA97/00473

EARLIER FILING DATE: 1997-07-04

EARLIER APPLICATION NUMBER: 60/021,272

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6  
LENGTH: 209  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-214-631-6

Query Match 6.8%; Score 168.5; DB 4; Length 209;  
Best Local Similarity 35.9%; Pred. No. 1e-06;  
Matches 42; Conservative 14; Mismatches 52; Indels 9; Gaps 3;

QY 33 YVMSANKRFOAE-----GGVLYPQIGDRDLDCPRARPGHSSPNVEFYLYVGGA 87  
DB 35 YVMSANKRFOAE-----GGVLYPQIGDRDLDCPRARPGHSSPNVEFYLYVGGA 93  
DB 94 GHSACDRHROGFKWECNRPAPGGLKFSKQQLFPFSLGFEFPFGHEYYISAT 150

RESULT 49  
US-08-442-248-4  
Sequence 4, Application US/08442248  
Patent No. 5759863  
GENERAL INFORMATION:  
APPLICANT: Caras, Ingrid W.  
TITLE OF INVENTION: AL-1 Neurotrophic Factor  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/442,248  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/330128  
FILING DATE: 27-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: 920C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8674  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 228 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-442-248-4

Query Match 6.8%; Score 167.5; DB 1; Length 228;  
Best Local Similarity 28.8%; Pred. No. 1.4e-06;  
Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;

QY 33 YVMSANKRFOAE-----GGVLYPQIGDRDLDCPRARPGHSSPNVEFYLYVGGA 85  
DB 34 YVMSANKRFOAE-----GGVLYPQIGDRDLDCPRARPGHSSPNVEFYLYVGGA 90  
QY 86 -----GAQRCEAPAPNLLTCDRDLDCPRARPGHSSPNVEFYLYVGGA 140  
DB 91 DHTSKGFRWECNRPAPGGLKFSKQQLFPFSLGFEFPFGHEYYISAT 141

QY 141 S---DGTREGESLGGVCLTRGKMLLRVQSPRGAVPRKPVSENMERDGAASLE 197  
DB 142 ALPDNGRS-----CLK--LKVPR-----PTNSCKMTIGHDVFPDNDVENSLE 186  
QY 198 PKEKLPDPTSNATSRGAGGLPPEPSMPAVAGAAGLALLLGA 243  
DB 187 PADDTV---HESASPSRG--ENNAQTPIPSRL-----LAILLFILA 223

RESULT 50  
US-08-440-815-4  
Sequence 4, Application US/08440815  
Patent No. 5798448  
GENERAL INFORMATION:  
APPLICANT: Caras, Ingrid W.  
TITLE OF INVENTION: AL-1 Neurotrophic Factor  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,815  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/330128  
FILING DATE: 27-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: 920C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-8674  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 228 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-440-815-4

Query Match 6.8%; Score 167.5; DB 1; Length 228;  
Best Local Similarity 28.8%; Pred. No. 1.4e-06;  
Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;

QY 33 YVMSANKRFOAE-----GGVLYPQIGDRDLDCPRARPGHSSPNVEFYLYVGGA 85  
DB 34 YVMSANKRFOAE-----GGVLYPQIGDRDLDCPRARPGHSSPNVEFYLYVGGA 90  
QY 86 -----GAQRCEAPAPNLLTCDRDLDCPRARPGHSSPNVEFYLYVGGA 140  
DB 91 DHTSKGFRWECNRPAPGGLKFSKQQLFPFSLGFEFPFGHEYYISAT 141  
QY 141 S---DGTREGESLGGVCLTRGKMLLRVQSPRGAVPRKPVSENMERDGAASLE 197  
DB 142 ALPDNGRS-----CLK--LKVPR-----PTNSCKMTIGHDVFPDNDVENSLE 186  
QY 198 PKEKLPDPTSNATSRGAGGLPPEPSMPAVAGAAGLALLLGA 243  
DB 187 PADDTV---HESASPSRG--ENNAQTPIPSRL-----LAILLFILA 223

Wed Feb 12 09:30:05 2003

us-10-021-121-2.rat

Page 23

Search completed: February 11, 2003, 12:06:22  
Job time : 21.5082 secs

---



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 11:48:59 ; Search time 33.195 Seconds  
(without alignments)  
1317.705 Million cell updates/sec

Title: US-10-021-121-2  
Perfect score: 2450  
Sequence: 1 MGPHSGPGGVGVGALLLLG.....TTLLRGRASVEAEGQHGPL 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	628.5	25.7	336	2	149766	hepatoma transmembr
2	623	25.4	346	2	149766	elk ligand - human
3	620.5	25.3	346	2	184743	hepatoma transmembr
4	604.5	24.7	345	2	148780	Strat/Ep1g2 proteas
5	599.5	24.5	345	2	158406	LEK-2 - rat
6	214.5	8.8	237	2	119914	hypothetical proteol
7	179	7.3	209	2	138849	LEK-3 - human
8	176	7.2	208	2	A54984	ELF-1 protein prec
9	175.5	7.2	213	2	JB0322	ephrin-A2 - human
10	170.5	7.0	228	2	A57084	regulative axon gui
11	169.5	6.9	201	2	138850	LEK-4 - human
12	167.5	6.8	228	2	158170	LEK-7 precursor -
13	166	6.8	205	2	A63377	B61 protein precu
14	159	6.5	180	2	S31216	collagen alpha 1(X
15	154.5	6.3	1049	1	CGH075	collagen alpha 1(X
16	153.5	6.3	1670	1	CGH075	collagen alpha 1(X
17	151.5	6.2	301	2	B31219	collagen alpha 3(I
18	149	6.1	325	2	T32248	hypothetical prote
19	149	6.1	569	2	S42886	collagen - alknot
20	148.5	6.1	316	2	T20497	hypothetical prote
21	148	6.0	921	2	S42617	collagen alpha 1(X
22	146.5	6.0	1315	2	A56101	collagen alpha 1(X
23	146.5	6.0	1492	2	A40333	collagen alpha 1(X
24	146.5	6.0	1774	2	B56101	collagen alpha 1(X
25	146	6.0	675	2	S20819	collagen alpha 3(I
26	145.5	5.9	675	2	T20819	hypothetical prote
27	145	5.9	674	2	S13301	collagen alpha 1(X
28	145	5.9	931	2	S13580	collagen alpha 1(X
29	144.5	5.9	438	2	S53787	collagen alpha cha

30	144	5.9	1027	2	S28774	collagen alpha cha
31	143	5.8	1747	2	A54121	collagen alpha-4 C
32	142.5	5.8	635	2	A57131	collagen alpha 2(V
33	142.5	5.8	743	1	S23779	collagen alpha 1(V
34	142.5	5.8	1496	1	CGH02V	collagen alpha 2(V
35	142	5.8	614	2	T33149	hypothetical prote
36	142	5.8	744	2	S15435	collagen alpha 1(V
37	142	5.8	1029	1	S21369	collagen alpha 2(V
38	142	5.8	1763	1	CGH01D	collagen alpha 2(V
39	141.5	5.8	1466	1	CGH07L	collagen alpha 1(I
40	141	5.8	319	2	T32250	hypothetical prote
41	141	5.8	744	1	A34246	collagen alpha 1(V
42	141	5.8	744	1	S23298	collagen alpha 1(V
43	140.5	5.7	305	2	T30165	collagen alpha 1(V
44	140	5.7	304	2	T22482	hypothetical prote
45	140	5.7	680	1	CGH01D	hypothetical prote
46	139.5	5.7	210	2	B44984	collagen alpha 1(X
47	139	5.7	940	2	JB0291	FB19 protein - hum
48	139	5.7	1024	2	S18251	collagen alpha 1(X
49	138.5	5.7	469	2	A24450	collagen alpha 2(V
50	138.5	5.7	1419	2	A41182	collagen alpha 1(I

ALIGNMENTS

RESULT 1  
149766  
hepatoma transmembrane kinase ligand - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
R/Bennett, B.D.; Zeigler, F.C.; Gu, O.; Fendly, B.; Goddard, A.D.; Gillett, N.; Matthews,  
Proc. Natl. Acad. Sci. U.S.A. 92, 1866-1870, 1995  
A/Title: Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kin  
A/Reference number: 149766; M01D:9519254; EMD:7534404  
A/Accession: 149766  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-336 <RES>  
A/Cross-references: GB:138847; NID:9769677; PIDN:AAC42052.1; PID:9769678  
C/Genetics:  
A/Genes: HTK

Query Match 25.7%; Score 628.5; DB 2; length 336;  
Best Local Similarity 41.8%; Pred. No. 3.1e-37;  
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

QY 14 GALLLGVLGVLGSLLEPYTNNSANKRPOABGVLYIQIGRDLDCPPRAPPFPHS 73  
DB 17 GLIMVLCRTAISRSIVLEPYNNSKFLPGQGLVLYPOIDDKDIIIPKV---DSKTV 73  
QY 74 PNEYFYKLVYVGAQGRCEAPPAPVLLTCDRPLDLRFTIKFOEYSPNLNGHEFRSH 133  
DB 17 GLIMVLCRTAISRSIVLEPYNNSKFLPGQGLVLYPOIDDKDIIIPKV---DSKTV 73  
QY 74 GQYEVYKVVYVDQADKCTIKENPILNCARPDDVFTIKFOEYSPNLNGHEFRSH 133  
DB 74 GQYEVYKVVYVDQADKCTIKENPILNCARPDDVFTIKFOEYSPNLNGHEFRSH 133  
QY 134 DYIITATSDTRPGLSLOGVCLTRGMKYLRLVGO--SPRGAVPRKVSSEMER-DR 190  
DB 134 DYIITATSDTRPGLSLOGVCLTRGMKYLRLVGO--SPRGAVPRKVSSEMER-DR 190  
QY 134 DYIITATSDTRPGLSLOGVCLTRGMKYLRLVGO--SPRGAVPRKVSSEMER-DR 190  
DB 134 DYIITATSDTRPGLSLOGVCLTRGMKYLRLVGO--SPRGAVPRKVSSEMER-DR 190  
QY 191 GAHSLSEPKENLPGDPTNATSRGAEGLPPSPMAVVAAGAALTLGLVAGAGAMC 250  
DB 194 GRSTTSFVKVPFGSGTNGNAGHSNNLLSSEVALPFGIAGSCITFVITITLVVLL 253  
QY 251 WRRRAKPSERHPGPGSGFGRGSLGLGGGGMGPPEAGPGLIALRGGAADPFCTH 310  
DB 254 KYRRRRKPSQHTTTLSTLSTLTPKRGANN---NGSEPSVLIPLR---TADSVFCBH 306  
QY 311 YEKVSGDYGPVYIVODGPOSPENITY 338  
DB 307 YEKVSGDYGPVYIVODGPOSPENITY 334

## RESULT 2

S46993

elk ligand - human

C/Species: Homo sapiens (man)

C/Date: 15-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 28-May-1999

C/Accession: S46993

R/Backmann, M.P.; Cerretti, D.P.; Baum, P.; vanden Bos, T.; James, L.; Farrah, T.; Kozl

EMBO J. 13, 3757-3762, 1994

A/Title: Molecular characterization of a family of ligands for eph-related tyrosine kin

A/Reference number: S46993; MUID:94349923; PMID:8070404

A/Accession: S46993

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-346 &lt;BEC&gt;

A/Cross-references: GB:U09304; NID:9538366; PIDN:AAA53093.1; PID:9538367

## Query Match

Best Local Similarity

Matches 143; Conservative

48; Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

25.4%; Score 623; DB 2; Length 346;

39.2%; Pred. No. 7.7e-37;

Mismatches 116; Indels 58; Gaps 9;

QY 14 GALLILGVLGVLGSLSPVYVWNSANKRFQAEAGVLYLPQIGDLDLCPARPPEHSS 73

Db 14 GLLVLCRTALSKSTVLEPIYVWNSNSKFLPGQGLVLYLPQIGDLDLCPVY--DSKTV 70

QY 74 PNYEFKLYLVGAAGRCCEAPPAVNLTLTCDRDLDFRTIKFOEYSPNLMGHEFRSH 133

Db 71 GQYEVKLYVMDKQADCTIKENTPLINCAKDQDIKFTIKFOEYSPNLMGHEFRSH 130

QY 134 DYIITATSDGREGIESLQGVCLTRGKVLRLVQ--SPRGAVPKPYSEMWER-DR 190

Db 131 DYIITATSDGREGIESLQGVCLTRGKVLRLVQ--SPRGAVPKPYSEMWER-DR 190

QY 191 GAHSLPEKXNLPDPTSNATSGABGLPPEPMPVAGAAGLALLGVAGAAGMC 250

Db 191 GRSSITSPVYKPNPSSSTDNAGSHGNNILGSEVALPAGIASCITFIITLVVLL 250

QY 251 WRRRAKPSBSRHPQPSFGGSLGAGGGGMPREAPPELGIATRGGAADPPCPH 310

Db 251 KRRRHRKPSQHTTSLTSLATPKRSNN--NGSEPDIILPLR--TADSVCFPH 303

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331

QY 311 YEKVSGDGHVYLYVODGPPSPNITY 338

Db 304 YEKVSGDGHVYLYVODGPPSPNITY 331





Best Local Similarity 29.9%; Pred. No. 5e-05;  
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGLSL-EPVYVNSANKRFQAEAGVLYLPQIGDRLDLCPRAPPGHSSPNFYLY 82  
DB 20 LKGGSLRRHVYVNSNPRL-DRGAVVELGINDLDVCPHYEPGPPEGP-ETFLY 76  
QY 83 LVGAQGRCEAP-PAPNLLTCDRPPDLRFPTIKFOEYSPNLMGHEFRSHHDYIATS 141  
DB 77 WDWMEYECQAEGRPAVYKWCSPFPGHVQSEKIQRTFPLSGFEFLPBTYYIIVP 136  
QY 142 DGTREGLESLOGVCLTRGMKVLRLVQSPRGCAVPRKPVSEMEPRDGAHSIEPKK 201  
DB 137 --TPR-----SSGQCL-----RLQVSVCKEKRSSESAHPV----- 164

QY 202 NLPGDPTNATS--RGAEGLPPSPMPAVAGAAGLALLL 240  
DB 165 ---GSPGESGTGMKRGDTPSP-----LCILL 189

RESULT 12  
LEK-7 precursor - human  
N:Alternate names: AL-1  
C:Species: Homo sapiens (man)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 29-Sep-1999  
C:Accession: L58170; G01812  
R:Winkler, J.W.; Moran, P.; Valverde, J.; Shih, A.; Yuan, J.Q.; Wong, S.C.; Tsai, S.P.; Neuron 14, 973-981, 1995  
A:Title: Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in reference number: L58170; MUID:95267434; PMID:7748564  
A:Reference number: L58170  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Accession: L58170  
A:Molecule type: mRNA  
A:Residues: 1-228 <RES>  
A:Cross-references: GB:S77167; NID:G914184; PID:G914185  
R:Koslosky, C.J.; Vandembos, T.; Park, L.S.; Cerretti, D.P.; Carpenter, M.K. submitted to the EMBL Data Library, May 1995  
A:Reference number: G08477  
A:Accession: G01812  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-228 <KOK>  
A:Cross-references: EMBL:U26403; NID:G1019430; PID:AA60377.1; PID:G1019431  
C:Genetic: GDB:EPUG7; AFL; LEK7  
A:Gene: GDB:EPUG7; AFL; LEK7  
A:Cross-references: GDB:568757; OMIM:601535  
A:Map position: 13q33-13q33  
C:Superfamily: axon guidance signal protein

Query Match 6.8%; Score 167.5; DB 2; Length 228;  
Best Local Similarity 28.8%; Pred. No. 8e-05;  
Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;

QY 33 VYVNSANKRFQAEAGVLYLPQIGDRLDLCPRAPPGHSSPNFYLYLVG----- 85  
DB 34 VYVNSNRPFC-RGYHIDVCLNDLDVCPHYEDSVEDKT-ERYULVWNTDGYAC 90  
QY 86 ----GAQGRCEAPAPNLLTCDRPPDLRFPTIKFOEYSPNLMGHEFRSHHDYIAT 140  
DB 91 DHTSKGFKEWECNRPSPV-----GFLKFSKFLFTFPFSGFEFFRGREFYISS 141  
QY 141 S---GSTRGLESLOGVCLTRGMKVLRLVQSPRGCAVPRKPVSEMEPRDGAHSLE 197  
DB 142 ALPDNGRRS-----CLK-LKVPR-----PNSCKMTIGVADRVDVNDKVENSL 186  
QY 198 PKEKNLPDPTNATSRGAEGLPPSPMPAVAGAAGLALLL 243  
DB 187 PADDTV---HESAEPSRG-ENNAQTPRIPSRL-----LAILLFLLA 223

RESULT 13  
A36377

B61 protein precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 29-Sep-1999  
C:Accession: A36377  
R:Holzman, L.B.; Marks, R.M.; Dixit, V.M.  
Mol. Cell. Biol. 10, 5830-5838, 1990  
A:Title: A novel immediate-early response gene of endothelium is induced by cytokines at A:Reference number: A36377; MUID:91042512; PMID:223719  
A:Accession: A36377  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-205 <HOL>  
A:Cross-references: GB:M57730; GB:M37476; NID:G179320; PID:AA58388.1; PID:G179321  
C:Superfamily: axon guidance signal protein

Query Match 6.8%; Score 166; DB 2; Length 205;  
Best Local Similarity 27.5%; Pred. No. 9.1e-05;  
Matches 46; Conservative 31; Mismatches 74; Indels 16; Gaps 5;

QY 18 LIGVLGVSLGSLPEVYVNSANKRFQAEAGVLYLPQIGDRLDLCPRAPPGHSSPN-- 75  
DB 8 LIGVCSLAADRHNTFVNSNPKFRNE-DYTIHQANDYVILICPHYB---DHSVADA 63  
QY 76 YEPYKLYVGAQGRCEAPAPNLLTCDRPL--DIRFTIKFOEYSPNLMGHEFRSH 132  
DB 64 MEQYIYLVHEHEYQLCQPSKQVWQCNRPSAKHGPKELEKQRTFPTLGRKFKSG 123  
QY 133 HDYIATSDGTREGLESLOGVCLTRGMKVLRLVQSPRGCAVPRK 179  
DB 124 HSYIYISKPIHQHEDR-----CLRLKTVSGKITHSPQAHVNFQE 163

RESULT 14  
S3126  
collagen alpha 1(X) chain precursor - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 13-Aug-1999  
C:Accession: S31216; S28807; S2215; S30127; I48299; S28397; S31830  
R:Kong, R.Y.C.; Kwan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; Eur. J. Biochem. 213, 99-111, 1993  
A:Title: Intron-exon structure, alternative use of promoter and expression of the mouse A:Reference number: S31216; MUID:93238750; PMID:8477738  
A:Accession: S31216  
A:Molecule type: DNA  
A:Residues: 1-680 <KON>  
A:Cross-references: EMBL:Z21610; NID:G49793; PIDN:CAA9736.1; PID:G49794  
R:Elima, K.; Berola, I.; Rosati, R.; Mettsaeranta, M.; Garofalo, S.; Peraelae, M.; de Crc Biochem. J. 289, 247-253, 1993  
A:Title: The mouse collagen X gene: complete nucleotide sequence, exon structure and ex A:Reference number: S28807; MUID:93143676; PMID:8424763  
A:Accession: S28807  
A:Molecule type: DNA  
A:Residues: 1-285, 'A', 287-680 <BLI>  
A:Cross-references: EMBL:X67348; NID:G50480; PIDN:CAA47763.1; PID:G50481  
R:Elima, K.; Mettsaeranta, M.; Kallio, J.; Peraelae, M.; Berola, I.; Garofalo, S.; de Crc Biochim. Biophys. Acta 1130, 78-80, 1992  
A:Title: Specific hybridization probes for mouse alpha-2(I) and alpha-1(X) collagen mR A:Reference number: S2215; MUID:92182017; PMID:1543751  
A:Accession: S2215  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 385-450, 'K', 452-627 <ELA>  
A:Cross-references: EMBL:X63013; NID:G49795; PIDN:CAA44741.1; PID:G49796  
R:Apfe, S.S.; Olsen, B.R.  
Matrix 13, 165-179, 1993  
A:Title: Characterization of the mouse type X collagen gene.  
A:Reference number: S30127; MUID:93261348; PMID:8492743  
A:Accession: S30127  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-12, 'F', 14-26, 'S', 28-247, 'L', 249-285, 'A', 287-305, 'F', 307-416, 'S', 418-499, 'L R:Apfe, S.S.; Seidin, M.F.; Hayashi, M.; Olsen, B.R.  
Eur. J. Biochem. 206, 217-224, 1992

A>Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse t  
 A/Reference number: 148299; MUID:92267014; PMID:1587271  
 A/Accession: 148299  
 A/Status: Preliminary; translated from GB/EMBL/DDJ  
 A/Molecule type: DNA  
 A/Residues: 35-247; L: 249-285; A: 287-305; F: 307-416; S: 418-499; L: 501-566; C: 566;  
 A/Cross-references: EMBL:X65121; NID:950482; PDB:CAA46237.1; PID:9667031  
 R/Summers, T.A.; Irtwin, M.H.; Mayne, R.; Balian, G.  
 J. Biol. Chem. 263, 581-587, 1988  
 A>Title: Monoclonal antibodies to type X collagen. Biosynthetic studies using an antibod  
 A/Reference number: 526397; MUID:88087150; PMID:2826450  
 A/Accession: 526397  
 A/Molecule type: protein  
 A/Residues: 'SDGYFSQ', 24-26; 'KQ' <SDM>  
 C/Genetics:  
 A/Gene: Col10a-1  
 A/Map position: 10  
 A/Intons: 51/3  
 C/Superfamily: collagen alpha 1(VIII) chain, complement C1q carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
 F:118/Domain: signal sequence #status predicted <SIG>  
 F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>  
 F:553-679/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 6.5%; Score 159; DB 2; Length 680;  
 Best Local Similarity 25.5%; Pred. No. 0.0011;  
 Matches 97; Conservative 30; Mismatches 108; Indels 146; Gaps 24;

QY 1 MGPHSGPGCV-RVYGLLLGLVGLSLPEYVNSANKRFOAEGGYLYPQIDRLD 59  
 DB 211 IGPF--GPSGVGRGNGFPQPGI-----KGRGFGENG----- 244  
 QY 60 LCCRAAPPSPHSPNYEYKLYVGAQRRCAPAPVLLTCRPPDILFTTKFOE 119  
 DB 245 ---PSAPPGPGPP-----GKQGR--EGKPKAIPSPGPGI----- 277  
 QY 120 YSPMLGHEFRSHDYIATSDT---REGLESLOGVCILRGKMYLLRVGQSPRGA 175  
 DB 278 --PGRKHPSGPG---LAPPPAPGFGKGLPLRG---QRG-----PAG-- 314  
 QY 176 VPKXVSEMERDGAHSLPEKENTLPDPTSNATSRGABPPLPPSPAVAGAAGL 235  
 DB 315 LPGAAPG---KGRGPGAG--HPGEPGLPGSP---GNMPQPGKGPNGHGPAGKGEI 364  
 QY 236 ALLLLGAGAGACMCRRAKRP---SESRHPC-----PGSRGRGSLGLGGGG 282  
 DB 365 G--LVGPAAGPPGA---GGAAGPPGLDGTGTPGPGINGKGNPGLPGQGDGVGTRPG 419  
 QY 283 M-----GPREAPGELGIALRGGAADPPCPHYEKVSGDYGHVY 323  
 DB 420 LRGVGVGAKVPGHNGEAGPR--GEPIGTR---GTPGPGVGVGFPGSKDPNP-- 472  
 QY 324 LVQDGP-----PQSP 334  
 DB 473 -GAPGPAIGATKGLNGPTGPP 492

RESULT 15  
 CGB07S  
 collagen alpha 1(III) chain - bovine  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 04-Dec-1986 #sequence, revision 04-Dec-1986 #extc change 07-May-1999  
 C/Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
 R/Fietzek, P.P.; Allmann, H.; Rauerberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
 A>Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen  
 A/Reference number: A02862; MUID:80026026; PMID:488906  
 A/Accession: A02862  
 A/Molecule type: protein  
 A/Residues: 1-242 <PIE>  
 A/Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A>Title: The covalent structure of calf skin type III collagen. II. The amino acid sequen

A/Reference number: A38001; MUID:80026027; PMID:488907  
 A/Accession: A38001  
 A/Molecule type: protein  
 A/Residues: 243-422 <DEW1>  
 R/Bentz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A>Title: The covalent structure of calf skin type III collagen. III. The amino acid sec  
 A/Reference number: A38002; MUID:80026028; PMID:488908  
 A/Accession: A38002  
 A/Molecule type: protein  
 A/Residues: 423-571 <BEN>  
 R/Liang, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
 A>Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequ  
 A/Reference number: A38003; MUID:80026029; PMID:488909  
 A/Accession: A38003  
 A/Molecule type: protein  
 A/Residues: 572-808 <LAN>  
 R/Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A>Title: The covalent structure of calf skin type III collagen. V. The amino acid sequ  
 A/Reference number: A38004; MUID:80026030; PMID:488910  
 A/Accession: A38004  
 A/Molecule type: protein  
 A/Residues: 809-947 <DEW2>  
 R/Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A>Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequ  
 A/Reference number: A38005; MUID:80026031; PMID:488911  
 A/Accession: A38005  
 A/Molecule type: protein  
 A/Residues: 948-1049 <ALU>  
 A/Experimental source: skin  
 R/Henkel, W.  
 Biochem. J. 318, 497-503, 1996  
 A>Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
 A/Reference number: S71946; MUID:96404897; PMID:8809038  
 A/Accession: S71946  
 A/Molecule type: protein  
 A/Residues: 87-106; 1017-1029; 1037-1049 <HEN>  
 C/Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are  
 C/Comment: The type III collagen molecule is a trimer of identical chains, linked to ea  
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol  
 F:11-1043/Product: collagen alpha 1(III) chain #status experimental <CAB>  
 F:11-14/Region: amino-terminal nonhelical telopeptide  
 F:15-1040/Region: helical  
 F:587-589/Region: cell attachment (R-G-D) motif  
 F:752-754/Region: cell attachment (R-G-D) motif  
 F:875-877/Region: cell attachment (R-G-D) motif  
 F:935-937/Region: cell attachment (R-G-D) motif  
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
 F:95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:107,950/Modified site: allysine (Lys) #status predicted  
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 6.3%; Score 154.5; DB 1; Length 1049;  
 Best Local Similarity 26.5%; Pred. No. 0.0036;  
 Matches 60; Conservative 9; Mismatches 74; Indels 83; Gaps 8;

QY 171 PRGAVPRKPPVSEMERDGA-----ASHLEPKEN 202  
 DB 668 PAGSGGAPGPGPGQGVKGRSGPGGPAAGPGRGPGPPSGNGNPGPSSGAPGMDG 747  
 QY 203 LPDPTSNAT-----SRGKGLPLPPSPPAVAGAAGLALLILGTVGA 245  
 DB 748 PPGPGSGAGAPSGPISGPKDGPPEGRGAPGQGPAGPLGIG-----LTGARGL 802  
 QY 246 GGAMCWRRAKRP-----SESRHPC-----GSGRGGSLGLGGGGGPREAPGE 292  
 DB 803 AGPPMPARSGPGRGKGNKGRGSGNGGNGGPPGPGPLPGLATGAPGADGNPS 862



QY 293 LGIARGS-----GAADPFCHYEKVGSDYGHPIVVDGPP 330  
 Db 863 DGLPGDGAFAKAGRGENGSGPAP-----GAPHP-----GPP 896

RESULT 16

CGHUB  
 collagen alpha 3(IV) chain precursor, long splice form - human  
 N:Alternate names: Goodpasture antigen, procollagen alpha 3(IV) chain long splice form  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text\_change 22-Jun-1999  
 C:Accession: A54763; A43928; A44043; A45971; A39786  
 R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Trygvaason, K.; Reeders, S.T.  
 J. Biol. Chem. 269, 23013-23017, 1994  
 A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
 A:Reference number: A54763; MUID:94364994; PMID:8083201  
 A:Accession: A54763  
 A:Molecule type: mRNA  
 A:Residues: 1-1670 <MAR>  
 A:Cross-references: GB:X80031; NID:G577563; PID:G577564  
 A:Experimental source: Kidney  
 R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
 J. Clin. Invest. 89, 592-601, 1992  
 A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al  
 A:Reference number: A43928; MUID:92147878; PMID:11737849  
 A:Accession: A43928  
 A:Molecule type: mRNA  
 A:Residues: 1331-1524, 'T', 1526-1670 <TUR>  
 A:Cross-references: GB:M61379  
 A:Experimental source: Kidney  
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 267, 19780-19784, 1992  
 A:Title: Exon/Intron structure of the human alpha 3(IV) gene encompassing the Goodpastur  
 A:Reference number: A44043; MUID:93015826; PMID:1400291  
 A:Accession: A44043  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1386-1670 <QUI>  
 A:Cross-references: GB:M92993; NID:G177895; PID:AAA210.1; PID:G177896  
 R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 269, 17358, 1994  
 A:Reference number: A44738; MUID:94274734; PMID:8006044  
 A:Contents: annotation; extram; correction to intronic sequence in A44043  
 R:Bernal, D.; Quinones, S.; Saus, J.  
 J. Biol. Chem. 268, 12090-12094, 1993  
 A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
 A:Reference number: A45971; MUID:93280184; PMID:8505332  
 A:Accession: A45971  
 A:Molecule type: mRNA  
 A:Status: nucleic acid sequence not shown  
 A:Residues: 1427-1444 <BBR>  
 A:Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly ident  
 R:Morrison, K.E.; Matiyama, M.; Yang-Peng, T.L.; Reeders, S.T.  
 Am. J. Hum. Genet. 49, 545-554, 1991  
 A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
 A:Reference number: A39786; MUID:91353570; PMID:1882840  
 A:Accession: A39786  
 A:Molecule type: mRNA  
 A:Residues: 1453-1593, 'A', 1595-1670 <WOR>  
 A:Cross-references: GB:SS7790; NID:9234418; PID:AA19637.1; PID:G234419  
 C:Comment: Proline and lysine at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope  
 C:Genetics:  
 A:Gene: GDB:COL4A3  
 A:Cross-references: GDB:128351; OMIM:120070  
 A:Map position: 2q36-q37  
 A:Introns: 1385/1, 1418/1, 1488/1, 1547/2, 1585/3, 1643/2 #status incomplete  
 A:Notes: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with  
 C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric

er associations in the interrupted helical domain (with disulfide and desmosine cross-1  
 A:Function:  
 A:Description: minor structural component of extracellular basement membrane in kidney;  
 C:Superfamily: collagen alpha 1(IV) chain  
 C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace  
 F:1-28/Domain: signal sequence #status predicted <SIG>  
 F:28-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>  
 F:28-42/Domain: amino-terminal nonhelical, NH1 <NH1>  
 F:43-138/Region: interrupted helical  
 F:791-793/Region: cell attachment (R-G-D) motif  
 F:996-998/Region: cell attachment (R-G-D) motif  
 F:1154-1156/Region: cell attachment (R-G-D) motif  
 F:1306-1308/Region: cell attachment (R-G-D) motif  
 F:1345-1347/Region: cell attachment (R-G-D) motif  
 F:1439-1439/Region: cell attachment (R-G-D) motif  
 F:1451-1551/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
 F:31,33,39,41,125,422,476,479,687,722,809,1387/Disulfide bonds: interchain #status pred  
 F:253/Binding site: carboxylate (asn) (covalent) #status predicted  
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted  
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 6.3%; Score 153.5; DB 1; Length 1670;  
 Best Local Similarity 27.0%; Pred. No. 0.0071;  
 Matches 108; Conservative 26; Mismatches 155; Indels 111; Gaps 25;

QY 2 GPHSGPGG---VVGALLILGVLGVLGSLFVYVNSANKR---FQABGVLYTP 52  
 Db 991 GPP---GPRGLGSGNGPGRGLGIGSGMGMGP---GSGKRGTLGFPGRAGPGLP 1044

QY 53 QI-----GDRDLICP-----RAPPGPHSPYERFYKYLVGAGQGRRC----- 93  
 Db 1045 GIGGLGDKKE---PGVSEGTREPPGPTGDC-----LPGMGKKGEMGQPPGPH 1093

QY 94 -----APPAPNL---LITCDPDLIRFT---IKFGYSPNIMGHE---FSGHDIYII 138  
 Db 1094 LGPAGBEGAPGSPGSPGLGPKPFGDLGFKGILKGLGPPGIRGPPGLGPFSGSPGPMGI 1153

QY 139 ATBDGREGLESIGGVCLTRGKVLRLVQSPRG---GAPPRKPSVEMEMERDGAHSL 196  
 Db 1154 RKGQG-RDGLPAPG---EKSGTGLRAPRPPGKPGQAGK-----GDRGA----- 1196

QY 197 EPKKNLPGDPTSNATSRGAEGLPPSPAYAGAAGLALLLGVAGAGAMCMRRRA 256  
 Db 1197 -PFPGLPG---RKAMDAGRPGPTGIEGFPGLGALIIPGGTNRGP---PGSRG 1248

QY 257 KPESRHPGP-----GSFRRGSLG-LGGGGGKMP---RAAEGBGLALRGCG 301  
 Db 1249 SPAPGPPGPGSHVIGIKDKKSMGHPGPKGPPGAGMGSPGRIGAGATPGLP---GP 1305

QY 302 AADPFCHYEKVGSDYGHPIVVDGP-----PGSPFNI 336  
 Db 1306 RQDPGF-QGPPGVKSGKNGPGLSGISGPPGPICPKPPGV 1344

RESULT 17

B31219  
 collagen 2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 30-Jun-1989 #sequence revision 30-Jun-1989 #text\_change 21-Jul-2000  
 C:Accession: B31219; T37289; T26033; T37288  
 R:Kramer, J.M.; Cox, G.N.; Hirsch, D.  
 Cell 30, 599-606, 1982  
 A:Title: Comparisons of the complete sequences of two collagen genes from Caenorhabditis  
 A:Reference number: A90825; MUID:83050944; PMID:7139711  
 A:Accession: B31219  
 A:Molecule type: DNA  
 A:Residues: 1-301 <RA>  
 A:Cross-references: GB:V00148; NID:96683; PIDN:CAA23464.1; PID:96684  
 A:Accession: T37289  
 A:Status: preliminary; translated from GB/EMBL/DBJ





A:Map position: X  
A:Introns: 56/3  
C:Superfamily: unassigned collagens

Query Match 6.1%; Score 148.5; DB 2; Length 316;  
Best Local Similarity 28.5%; Pred. No. 0.0025;  
Matches 49; Conservative 12; Mismatches 78; Indels 33; Gaps 6;  
166 GQSPGAVPRFVSEMPERDRGAHSLPEKENTLPDPP-----TSNATSTGA 216  
Db 144 GASGKGAAPCFSTPTTPCQ-PCPAGPPEPPDGTGPGGPGPAGSPAGSPGPGAP 202  
Qy 217 EGPPLPPSPNPAVAGAGLALLLLVAGAGAMCWRBRAPKPSERHGP-GSFGRGSL 275  
Db 203 PGPAPPGNDGPGQPGPGG- QDGAASAG-----EAGPGAPGPGAPGAPGPGQS 253  
Qy 276 GGGGGGKMPR-----EAPRGLGLRGGADPPFCYKAVSG 316  
Db 254 GSGSAGGPGKGP PGPAGQPGSDGNPGTAGPPNPGGEGEKICPKYCAIDG 305

## RESULT 21

S42617  
collagen alpha 1(IX) chain - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 25-Dec-1994 #sequence\_revision 19-Apr-1996 #text\_change 20-Sep-1999  
C/Accession: S42617  
R/Rokos, I.; Muragaki, Y.; Warman, M.; Olsen, B.R.  
Matrix Biol. 14, 1-8, 1994  
A/Title: Assembly and sequencing of a cDNA covering the entire mouse alpha-1(IX) collagen  
A/Reference number: S42617; MUID:94340199; PMID:8061915  
A/Status: preliminary  
A/Accession: S42617  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-921 <ROR>  
A/Cross-references: GB:112215; GB:119396; NID:9192675; PTDN:AAA21834.1; PTD:9192676  
C:Superfamily: unassigned collagens

Query Match 6.0%; Score 148; DB 2; Length 921;  
Best Local Similarity 24.0%; Pred. No. 0.009;  
Matches 87; Conservative 23; Mismatches 130; Indels 122; Gaps 17;

Qy 2 GPHHGGPGVVRGALLLGLVGLSLPEPYNNANKRQAEQGYLYPQIGDRDL 61  
Db 391 GPP--GPPPPS-----GTIGFHDG-----DPL 410  
Qy 62 CPRAPPSPHSSPNYEFKLYLVG-----GAQRRCAPAPNLLTCDRPDLIRFTIK 116  
Db 411 CPNSCPSPRSQVPG-----LPGRKRGKAKGEIGF----- 440  
Qy 117 FOEYSPNIMGHEFRSHNDYIATSDGTREGLESIQG--GVCLTRGMKVLRLVQSPRG 174  
Db 441 ----PGRQGHGSEBDQGEVVDGPPGQGLRGITGIVGDGEGKAGFGESEPPQ 495  
Qy 175 AVP-----RKPVSEMPMERDRGAHSLPEKENTLPDPP-----TSNATSTGA 216  
Db 466 GTPGAAGDGGQGTGPGTGTGPEEDRGI-----QGSRTGIGSGEPKQDGTLPSEVDRDGI 550  
Qy 225 MPVAVGAAG-----GLALP--LLVAGAGAMCWRBRAPKPSERHGP- 275  
Db 551 NPGTSGEAGKPPGPDVGAIGLPGVPGI-PGAKGVAGEKNTGAPKPGQLSSGKPPGQ 610  
Qy 266 -PGSGRGGSLG--GGGGGMPREAE--PGLGLRLRG--GAADPPCPHYEKVSGDY 315  
Db 611 PPGVEVPPRPLPGSRGPGVGGSGPGI-PGKLGVSGPGLPGLPGLPGMKGDRGVFG 670  
Qy 320 HP 321  
Db 671 EP 672

## RESULT 22

A56101

collagen alpha 1(XVIII) chain precursor, short splice form - mouse  
N/Context: endostatin  
C/Species: Mus musculus (house mouse)  
C/Date: 03-Oct-1995 #sequence\_revision 08-May-1998 #text\_change 31-Mar-2000  
C/Accession: A56101; A56371; S72450; S65595  
R/Rehn, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995

A/Title: Identification of three N-terminal ends of type XVIII collagen chains and their homologous to rat and Drosophila fibrillar proteins.  
A/Reference number: A56101; MUID:95181468; PMID:7876242  
A/Accession: A56101

A/Molecule type: mRNA

A/Residues: 1-103 <REH2>

A/Cross-references: GB:111636; NID:9618427; PTDN:AA052176.1; PTD:9618428

R/Rehn, M.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994

A/Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous  
A/Reference number: A56371; MUID:94240112; PMID:8183894

A/Accession: A56371

A/Molecule type: mRNA

A/Residues: 1-928 <REH2>

A/Cross-references: GB:1116899; NID:9404754; PTDN:AAA37434.1; PTD:9553894

R/Oh, S.F.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
Submitted to the EMBL Data Library, August 1993

A/Reference number: S72450

A/Accession: S72450

A/Molecule type: mRNA

A/Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>

A/Cross-references: EMBL:U22445; NID:9348968; PTDN:AAA15787.1; PTD:9511298

R/Oh, S.F.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A/Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa

A/Reference number: A56370; MUID:94240111; PMID:8183893

A/Accession: S65595

A/Molecule type: mRNA

A/Residues: 28-1315 <OHM>

A/Cross-references: EMBL:U22445

C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
labeled and subsequently O-glycosylated.

C/Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per-

ay be useful in treating solid tumors.

C/Genetics:

A/Genes: MGI:Col18a1

A/Cross-references: MGI:71175

A/Map position: 10:41.0

C/Suprafamily: unassigned collagens

C/Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc

P/-25/Domain: signal sequence #status predicted <SIG>

F/4-235/Region: thrombospondin amino-terminal similarity

F/26-315/Region: collagen alpha 1(XVIII) chain, short splice form #status predicted <P

F/327-353/Domain: collagenous #status predicted <CO1>

F/462-583/Domain: collagenous #status predicted <CO2>

F/607-689/Domain: collagenous #status predicted <CO3>

F/704-745/Domain: collagenous #status predicted <CO4>

F/759-831/Domain: collagenous #status predicted <CO5>

F/842-874/Domain: collagenous #status predicted <CO6>

F/887-910/Domain: collagenous #status predicted <CO7>

F/892-894/Region: cell attachment (R-G-D) motif

F/918-969/Domain: collagenous #status predicted <CO8>

F/983-1000/Domain: collagenous #status predicted <CO9>

F/1132-1315/Region: collagenous #status predicted <CO10>

F/1133-1315/Region: multiplexin collagen carboxyl-terminal similarity

F/126-488/Binding site: carbohydrate (Asn) (covalent) #status predicted

F/172-228/Disulfide bonds: #status predicted

F/240-245/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

F/451-454/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 6.0%; Score 146.5; DB 2; Length 1315;  
Best Local Similarity 23.7%; Pred. No. 0.017;  
Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

```

QY 2 GPHSGGGRVAGALLLGLG-LVSGLSLEPYMNSAKRFOAGGVLYLFOIGDRDL 60
Db 339 GPR--GPAGQ-----GPAGVYVOSVNSQPV----- 362
QY 61 LCPRAR-PPEPHSSPNVEFYKLVGAQGRCE----- 93
Db 363 --FGAQGPFPQGPGRKD-----GTPRGDEPDGPDGDRPDGTGQGFPGTGDVG 412
QY 94 -----APPANILLTCDBPDLDLRTIKFOETSPILMHERSHHDYITAT 140
Db 413 PKRGKDPGIGPPGPPGP-----PPPPPSFRDDKLTFTIDME 449
QY 141 SDGTRELESLQGVCLTRGMKYLRYVQSPRGCAVPRKP--VSEPMRDR- GAASLLE 197
Db 450 GSGPESGIEILRG-----PRGPPGPPGPVGLPGSGRFGINGSYA 492
QY 198 PKENLPDPTSNATSGAEP--LP-PSPMPAVAGAAGLALLLGVAGAGAMCMRRR 254
Db 493 PGPAGLPGVP-----GKEGPPGPPGPPGPGRPKGEP-----PGVAGQKGSV----- 535
QY 255 RAKPSESRHGP-GSFRGSLGLGGGGMGPRAEPBELGLLRGG--GAADPPPCPHY 311
Db 536 ----GDVGLPGKSGKDLGPIMPGSGLA--GSPGVGPPGPPGPPGPPGPAAGP 588
QY 312 EKVSQDYGHFVYIQ-----DGPPQSP 333
Db 589 DMEGS-GTPLMTTARSSDGLQGPSP 615

RESULT 23
A4033
collagen alpha 1'(II) chain precursor - African clawed frog
C/Species: Xenopus laevis (African clawed frog)
C/Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #ext_change 16-Jul-1999
C/Accession: A40333
R/Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A/Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis en
A/Reference number: A40333; PMID:92011898; PMID:1918153
A/Accession: A40333
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1492 <SDA>
A/Cross-references: GB:M63596
A/Note: this sequence is presented as substitutions relative to another sequence in a fi
as they replace the appropriate interpretation of the sequence figure was reconstructed
C/Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology?
C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 6.0%; Score 146.5; DB 2; Length 1492;
Best Local Similarity 31.1%; Pred. No. 0.02; Indels 33; Gaps 8;
Matches 56; Conservative 11; Mismatches 80;

QY 171 PRGGAVERKVESEMPERDRG-AAHSLPEKENTLPPDPTSNATSGAEGPLPPSPMAVA 229
Db 123 PRGPPGQSGEGCGPRGERGDKETGAPGRGRDEP---GTPNPAPAGPPGPPCP 178
QY 230 GAAGGATALLLG--VAGAGAMCMRRR-----RAKSESHRPPSGF-GRGSLGIG 278
Db 179 GLGGRNPLAQTGFDEKAGQMGVMOGPMGPRGPPGSPAPRQGPQGNPGRGSGS 238
QY 279 GGGG-MGPR-----EAPFELGIALRG-----GAADPPPCPHYEKVSGDYGHP 321
Db 239 GAGGPMGRPRPPGSGRPPGDDDEAKRKGSGERGPGRGQARGFPTPLGLVKGHRGVP 298

RESULT 24
B5101
collagen alpha 1(XVII) chain precursor, long splice form - mouse
N/Conting: collagen alpha 1(XVII) chain precursor, medium splice form; endostatin
C/Species: Mus musculus (house mouse)

```

```

C/Date: 03-Oct-1995 #sequence_revision 08-May-1998 #ext_change 31-Mar-2000
C/Accession: B56101; C56101; S72450; S65595; P06775; A54072; A58816
R/Rein, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A/Title: Identification of three N-terminal ends of type XVII collagen chains and tiss
rif homologue to rat and Drosophila fibrinized proteins
A/Reference number: A56101; PMID:95181468; PMID:7876242
A/Accession: B56101
A/Molecule type: mRNA
A/Residues: 1-562 <REH1>
A/Cross-references: GB:U11637; NID:G618429; PIDN:AAC52179.1; PID:G618430
A/Experimental source: splice form clone PE17.24
A/Accession: C56101
A/Molecule type: mRNA
A/Residues: 1-239,487-562 <REH2>
A/Cross-references: GB:U11637; NID:G618429
A/Experimental source: splice form clones PE8.1, PE19, PE15.2
R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A/Reference number: S72450
A/Accession: S72450
A/Molecule type: mRNA
A/Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-1
A/Cross-references: EMBL:L22545; NID:G348968; PIDN:AA19787.1; PID:G511298
R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A/Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xa
A/Reference number: A58370; NID:94240111; PMID:8183893
A/Accession: S65595
A/Molecule type: mRNA
A/Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A/Cross-references: EMBL:L22545
R/Abe, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Niinomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A/Title: Identification of a novel collagen chain represented by extensive interruption
A/Reference number: P06775; NID:94059075; PMID:8240330
A/Accession: P06775
A/Molecule type: mRNA
A/Residues: 635-1774 <ABE>
R/Rein, M.; Hinkka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A/Title: Primary structure of the alpha1 chain of mouse type XVII collagen, partial st
collagen chain.
A/Reference number: A54072; NID:94245707; PMID:8186673
A/Accession: A54072
A/Molecule type: DNA; mRNA
A/Residues: 1293-1403, 'R', 1405-1774 <REH3>
A/Cross-references: GB:U03114; NID:G487733; PIDN:AA20657.1; PID:G487734
R/O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B
Cell 88, 277-285, 1997
A/Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A/Reference number: A58816; NID:97160848; PMID:9008168
A/Accession: A58816
A/Molecule type: protein
A/Residues: 1591-1610 <ORE>
A/Experimental source: hemangioendothelium cells
A/Note: inhibits endothelial cell proliferation
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C/Comment: The different splice forms of collagen alpha 1(XVII) may be involved in per
C/Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of u
ay be useful in treating solid tumors.
C/Genetics:
A/Gene: MG1:Coll18a1
A/Cross-references: MG1:71175
A/Map position: 10:41.0
A/Intons: 1295/3; 1310/1; 1331/1; 1345/3; 1368/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599
A/Note: the 1st of introns is incomplete
C/Superfamily: unassigned collagens
C/Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F:1-1774/Product: collagen alpha 1(XVII) chain precursor, long splice form #status pre
F:1-239,487-1774/Product: collagen alpha 1(XVII) chain precursor, medium splice form #
F:1-24/Domain: signal sequence #status predicted <SIG>

```

F1361-486/Region: frizzled similarity  
 F1786-812/Domain: collagenous #status predicted <CO01>  
 F1823-896/Domain: collagenous #status predicted <CO02>  
 F1921-1042/Domain: collagenous #status predicted <CO03>  
 F11066-1148/Domain: collagenous #status predicted <CO04>  
 F11163-1204/Domain: collagenous #status predicted <CO05>  
 F11218-1290/Domain: collagenous #status predicted <CO06>  
 F11301-1333/Domain: collagenous #status predicted <CO07>  
 F11346-1369/Domain: collagenous #status predicted <CO08>  
 F11351-1353/Region: cell attachment (R-G-D) motif  
 F11377-1428/Domain: collagenous #status predicted <CO09>  
 F11442-1459/Domain: collagenous #status predicted <CO10>  
 F11591-1774/Product: endostatin #status predicted <EST>  
 F11598-1774/Region: multidomain collagen carboxyl-terminal similarity  
 F1354-361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F1699-704,1716/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted  
 F1910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 6.0%; Score 146.5; DB 2; Length 1774;  
 Best Local Similarity 23.7%; Pred. No. 0.024;  
 Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

QY 2 GPHSGPGVAVGALLLVGVLGVLGSLSPVYWNANKRFAAGGVLYVPIGDKL 60  
 Db 798 GPF--GPAQPQ-----GPAQVQSPNSQPV----- 821  
 QY 61 LCPRAR-PPGPHSPNYEFYKLYLVGAQGRCE----- 93  
 Db 822 --FGAQGPGPGPGPKGK-----GTPGADGPDGPDGPDGPDGPDGPDG 871  
 QY 94 -----APPANLLITCDRPDLIRFTTKQVSPNLMGHEFRSHDYITAT 140  
 Db 872 PKGEGKDPGIGRPGPGP-----PGPSPSPRQDKLTFIDME 908  
 QY 141 SDGTREGSEISOGVCLRGKMKVLRVQSPRGAGVPRK--VSEMEMRDR-GAHSLE 197  
 Db 909 GSGFSGDIESTLRG-----PRGPPGPGPGVGLGPEPRFGINSYA 951  
 QY 198 PGKENTPGDPTSNATSRGAEGR--LP-PPSMFAVAGAAGIALLLLVAGAAGMCRWR 254  
 Db 952 PGPAQLPQV-----GKEGPPGPGPGPGPGPKGEGP-----PGVAGKSGV----- 994  
 QY 255 RAKBESHSRPP-GSGFGGSLGCGGGGMPREARPEGLIALRG--GAADPPFCPHY 311  
 Db 995 ---GDVGI PGPCKSGKDLGPIGMFGKSGLA---GSPGPVGP PGP PGP PGP PGP 1047  
 QY 312 EKVSQDYGHPIVQ-----DGPQSP 333  
 Db 1048 DDMGSG-GIPMTTARASDGIQGPPEP 1074

RESULT 25  
 S20819  
 collagen alpha 3 (IX) chain precursor - chicken  
 C/Species: Gallus gallus (chicken)  
 C/Date: 22-Nov-1993 #sequence revision 09-Mar-1996 #text change 20-Sep-1999  
 C/Accession: S20819; S22429; S22918; S22338; C18856; S22241  
 R/Brewton, R.G.; Ouspenskaya, M.V.; van der Rest, M.; Mayne, R.  
 Eur. J. Biochem. 205, 443-449, 1992  
 A/Title: Cloning of the chicken alpha3 (IX) collagen chain completes the primary structure  
 A/Reference number: S20819; MUID:92241276; PMID:1572350  
 A/Accession: S20819  
 A/Molecule type: mRNA  
 A/Residues: 1-675 <BR>  
 A/Cross-references: EMBL:X64712; NID:963316; PIDN:CAA45967.1; PID:963317  
 R/Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;  
 J. Biol. Chem. 267, 10070-10076, 1992  
 A/Title: Cloning and developmental expression of the alpha3 chain of chicken type IX col  
 A/Reference number: S22429; MUID:92250566; PMID:1577778  
 A/Accession: S22429  
 A/Molecule type: mRNA  
 A/Residues: 1-195, 'G', 197-675 <HA1>  
 A/Cross-references: EMBL:M31179

A/Note: 353-Arg, 386-Ileu and 548-Arg were also found  
 R/Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.  
 submitted to the EMBL Data Library, February 1992  
 A/Description: Cloning and developmental expression of the alpha 3 chain of chicken type  
 A/Reference number: S22918  
 A/Accession: S22918  
 A/Molecule type: mRNA  
 A/Residues: 1-195, 'G', 197-405, 'S', 407-675 <HA2>  
 A/Cross-references: EMBL:M31179; NID:9211040; PIDN:AA59960.1; PID:9211041  
 R/Mayne, R.; van der Rest, M.; Nimmiya, Y.; Olsen, B.R.  
 Ann. N. Y. Acad. Sci. 460, 38-46, 1985  
 A/Title: The structure of type IX collagen.  
 A/Reference number: S22238; MUID:86185164; PMID:386958  
 A/Accession: S22238  
 A/Molecule type: protein  
 A/Residues: 540-548 <HA3>  
 R/Nimmiya, Y.; van der Rest, M.; Mayne, R.; Lozano, G.; Olsen, B.R.  
 Biochemistry 24, 4223-4229, 1985  
 A/Title: Construction and characterization of cDNA encoding the alpha2 chain of chicken  
 A/Reference number: A18856; MUID:86026268; PMID:2996593  
 A/Accession: C18856  
 A/Molecule type: protein  
 A/Residues: 540-558 <NIN>  
 R/Shimokomaki, M.; Wright, D.W.; Irwin, M.H.; van der Rest, M.; Mayne, R.  
 Ann. N. Y. Acad. Sci. 580, 1-7, 1990  
 A/Title: The structure and macromolecular organization of type IX collagen in cartilage  
 A/Reference number: S22241; MUID:90247791; PMID:2186687  
 A/Accession: S22241  
 A/Molecule type: protein  
 A/Residues: 135, 'E', 137-187, 'X', 189-191 <SH1>  
 C/Superfamily: unassigned collagens  
 C/Keywords: coiled coil; connective tissue; disulfide bond; extracellular matrix; heter  
 F1-21/Domain: signal sequence #status predicted <SIG>  
 F1-22-24/Product: collagen alpha 3 (IX) chain #status predicted <MAT>  
 F1-22-24/Domain: non-collagenous NC4 #status predicted <NC4>  
 F1-25-161/Domain: collagenous COL3 #status predicted <COL3>  
 F1-162-176/Domain: non-collagenous NC3 #status predicted <NC3>  
 F1-177-515/Domain: collagenous COL2 #status predicted <COL2>  
 F1-516-546/Domain: non-collagenous NC2 #status predicted <NC2>  
 F1-547-657/Domain: collagenous COL1 #status predicted <COL1>  
 F1-658-675/Domain: non-collagenous NC1 #status predicted <NC1>  
 F1-22/Modified site: pyrrolidone carboxylic acid (Glu) (in mature form) #status predicted  
 F1-37,144,146,149,152,155,179,182,185,552,558/Modified site: hydroxyproline (Pro) #statu  
 F1-170,174,525,658,663/Disulfide bonds: interchain #status predicted

Query Match 6.0%; Score 146; DB 2; Length 675;  
 Best Local Similarity 28.8%; Pred. No. 0.0089;  
 Matches 64; Conservative 9; Mismatches 71; Indels 78; Gaps 12;

QY 166 RVQSPFGAVPRKVSSEMPERDRGAHSLPGKENTPGDPTSNATSRGAEPLPPSM 225  
 Db 23 RVG--PQPPGPGPGPGPSGKXGIDG-----EPGSGLP PGP-----GPGAGRGK 65  
 QY 226 PAVAGAAG--GIALLLLVAGAAGMCRWRRAKPSRHPG-----PGSFGRG- 272  
 Db 66 PAAAGAGLPGIP-----GVDGLTGT-----DGPSPGPGPDRCALGPAGP PGP PGP 115  
 QY 273 -----GSLGIGGGGCV-----GPR-----EAPSELGIALRG----- 300  
 Db 116 PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP 175  
 QY 301 -GAADPPFCPHYEXKSGDYGHPIVQDP-----PQSPPI 336  
 Db 176 PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP PGP 217

RESULT 26  
 T20906  
 hypothetical protein F147.1 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 21-Jan-2000  
 C/Accession: T20906  
 R/McMurry, A.



A:Molecule type: DNA  
 A:Residues: 1-4,233-248, 'T', 250-267 <OLS>  
 A:Cross-references: EMBL:M32135  
 R:Diab, M.; Wu, J.-J.; Eyre, D.R.  
 Biochem. J. 314, 327-332, 1996  
 A:Title: Collagen type IX from human cartilage: a structural profile of intermolecular  
 A:Reference number: S64673; MUID:96195147; PMID:8660302  
 A:Accession: S74294  
 A:Molecule type: protein  
 A:Residues: 405-417 <DIA>  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C:Genetics:  
 A:Gene: GDB:COL9A1  
 A:Cross-references: GDB:119794; OMIM:120210  
 A:Map position: 6q12-6q14  
 A:Intons: 5/2; 232/3; 260/3; 267/3  
 C:Complex: type IX collagen may be a heterotrimer of one alpha 1 (IX) chain, one alpha 2  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer associated with typ  
 C:Superfamily: unassigned collagens  
 C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; heter  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-931/Product: collagen alpha 1 (IX) chain, long splice form #status predicted <MAT>  
 F:24-268/Domain: non-collagenous NC4 #status predicted <NC4>  
 F:269-405/Domain: non-collagenous COL3 #status predicted <COL3>  
 F:406-417/Domain: non-collagenous NC3 #status predicted <NC3>  
 F:418-786/Domain: non-collagenous COL2 #status predicted <COL2>  
 F:787-901/Domain: non-collagenous NC2 #status predicted <NC2>  
 F:902-931/Domain: non-collagenous COL1 #status predicted <COL1>  
 F:171/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.9%; Score 145; DB 2; Length 931;  
 Best Local Similarity 22.3%; Pred. No. 0.015;  
 Matches 84; Conservative 21; Mismatches 120; Indels 152; Gaps 13;

QY 2 GPPHSGPGVAVGALLLGLVLSGLSEPVVNSANKRQAGVYVLPQTGDRDL 61  
 DB 391 GPP- GPPGPR-----GTIGPHDG-----DPL 410  
 QY 62 CPRAPPGPHSSPNVEFYKLVG-----GAQGRCEAPAPNLLTCDRDLRFTIK 116  
 DB 411 CPNACPPGSRSGVPG-----LPGWRGHKAKGEIG----- 440  
 QY 117 FOEYSPNMGHEFSHHDYIATSDGRELSLOGVCLTRGMKYLAVGSPRGAV 176  
 DB 441 -----PGRQGH-----KKEEDQG-----ELGEVGAQGGPPGAQ 468  
 QY 177 PRKPVSEMPERDRGAASLT--EPKENTLPDPTSNAT-----SRGAGPLPP 223  
 DB 469 GLRGITGLVGDKKKRGARGLDGERPGQLPGAPDQQRPGRPGAKRGABGARIP 528  
 QY 224 SMPAVAGAGLALLLGVAGAGGACMCRRRRAPKPSRHFG-----PGRGSGSLG 276  
 DB 529 GLPPEKPGDTG-----LPGVDRDIPGMWPGKPGKPGPDAGLQGLGVGIPGAKG 583  
 QY 277 LGGGGMGPREAPGEGELALRG-----GANDPFCPHYEVAVSG 316  
 DB 584 VAGKRGSGAGKRGKQMNKSGKPGQDQGPGEVGPGRPGQLPGRGELGVSGPGLPKLG 643  
 QY 317 DYGHVYIVDDGPPQSP 333  
 DB 644 SLGSPGLPGLPGLP 660

RESULT 29  
 S53787  
 collagen alpha chain - Paravivnella grasselet (fragments)  
 C:Species: Paravivnella grasselet  
 C:Date: 19-Mar-1997 #sequence\_revistion 18-Jul-1997 #text\_change 07-May-1999  
 C:Accession: S53787

R:Gall, F.; Mann, K.; Wiedemann, H.; Engel, J.; Timpl, R.  
 J. Mol. Biol. 246, 284-294, 1995  
 A:Title: Structural comparison of cuticle and interstitial collagens from annelids liivi  
 A:Reference number: S53786; MUID:95173973; PMID:7869380  
 A:Accession: S53787  
 A:Molecule type: protein  
 A:Residues: 1-90,91-254,255-304,305-374,375-438 <GAI>  
 C:Superfamily: unassigned collagens  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein

Query Match 5.9%; Score 144.5; DB 2; Length 438;  
 Best Local Similarity 26.5%; Pred. No. 0.007;  
 Matches 101; Conservative 15; Mismatches 124; Indels 141; Gaps 22;

QY 2 GPP-----HSGPGGVAVGALLLGLVLSGLSEPVVNSANKRQ 43  
 DB 105 GPPGLTGSXGPAG-PXGANGLPQTQGFAGAGPXSPPGPRGGLGGRP---GEAGKRGQ 160  
 QY 44 AEGGYLYPQ-----IGDRLDLCTRAAP-----GPHSPNVEFYKLVG 85  
 DB 161 RGVPGALGPPGQGERGLPGER---GLPGRSGPGGARGGPRGSGPP-----G 207  
 QY 86 GAQG--RRCAPAPNLLTCDRPLDLRFTIKFOEYSPNMGHEFSHHDYIATSDG 143  
 DB 208 GPQGSQGRAGCTGAGGLRAGAPRP-----SDG 235  
 QY 144 TRGLESLOG-----GVCLTRGMKYLAVGSPRGAVPRKPVSEMPERDRGAASLEP 198  
 DB 236 -BEGRKSSGXPKDQK-----PPQDGPFR-----LPCRGAAGS--N 269  
 QY 199 GKENTLPDPTSNATSRGAGPLPPSPMPAVAGAG--GLALLLGVAGAGACMCRRRRA 256  
 DB 270 GEDGQNGK-----GGKGVGPPGICGLPGLPEERAKGELGLPEERK--PGRG 319  
 QY 257 KPESRHPP-----PSPFRGSGSLGAGGCGMCPRAEPGELGALRGGAADPEPCPYE 312  
 DB 320 KQGESGHPTAGLPDPGPGGV--KGRGDHGR-GEFGSPG---SXGERPPGXFXPQ 372  
 QY 313 KYSGVGHVYIVDDGPPQSP 333  
 DB 373 GLRGDRKNTGERGEPPAPGPR 393

RESULT 30  
 S28774  
 collagen alpha chain - tube worm (Riftia pachytila) (fragment)  
 C:Species: Riftia pachytila  
 C:Date: 22-Nov-1993 #sequence\_revistion 09-Mar-1996 #text\_change 23-May-1997  
 C:Accession: S28774; S22915; S17581  
 R:Mann, K.; Gall, F.; Timpl, R.  
 Eur. J. Biochem. 210, 839-847, 1992  
 A:Title: Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen fr  
 A:Reference number: S28774; MUID:93130909; PMID:1483468  
 A:Accession: S28774  
 A:Molecule type: protein  
 A:Residues: 1-95, 'X', 97-107, 'X', 109-191, 'X', 193-260, 'X', 262-278, 'X', 280-572, 'X', 574-611,  
 A>Note: we have shown the unidentified residues as lys forming glycosylated 5-hydroxylys  
 R:Mann, K.; Gall, F.; Timpl, R.  
 submitted to the Protein Sequence Database, July 1992  
 A:Description: Amino acid sequence and cell adhesion activity of a fibril-forming collac  
 A:Reference number: S22915  
 A:Accession: S22915  
 A:Molecule type: protein  
 A:Residues: 1-95, 'X', 97-107, 'X', 109-191, 'X', 193-260, 'X', 262-278, 'X', 280-572, 'X', 574-611,  
 A>Note: 903-proline modified to 4-hydroxyproline was also found  
 R:Gall, F.; Wiedemann, H.; Mann, K.; Kuehn, K.; Timpl, R.; Engel, J.  
 J. Mol. Biol. 221, 209-223, 1991  
 A:Title: Molecular characterization of cuticle and interstitial collagens from worms col  
 A:Reference number: S17581; MUID:92015209; PMID:11920405  
 A:Accession: S17581  
 A:Molecule type: protein  
 A:Residues: 8-45,525-545, 'X', 547-566, 'X', 568-572, 'X', 574-611, 'X', 613-618, 'X', 611-882 <GA  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (

h 3- and 4-hydroxylated in the X-position. Lysines are 5-hydroxylated and subsequently a  
 C/Complex: homotrimer  
 C/Superfamily: unassigned collagens  
 C/Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; homotrimer; h  
 F/1-12/Domain: amino-terminal telopeptide (fragment) <NTS>  
 F/13-1023/Domain: collagenous #status experimental <COL>  
 F/1024-1027/Domain: carboxyl-terminal telopeptide (fragment) <CTB>  
 F/21-24,123,243,473,726,285,291,303,348,381,621,645/Modified site: 4-hydroxyproline (Pro  
 F/27,39,54,72,90,93,128,150,162,165,174,177,180,207,216,219,228,237,249,255,306,312,321,  
 711,714,717,723,744,759,774,783,792,816,843,849,855,861,867,888,894,915,945,954,963,966  
 F/53,161,165,416,551,647,773,815,1010,1013,1016,1019/Modified site: 3-hydroxyproline (Pro  
 F/96,108,192,261,279,573,612,657,738,765,810,927,936/Modified site: 5-hydroxylysine (Lys  
 F/96,108,192,261,279,573,612,657,738,765,810,927,936/Binding site: carboxylate (Lys) (C  
 F/103,342,546,567,933/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F/351,933/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 5.9%; Score 144; DB 2; Length 1027;  
 Best Local Similarity 22.2%; Pred. No. 0.019;  
 Matches 82; Conservative 24; Mismatches 113; Indels 150; Gaps 17;

QY 46 GGYLYYQIGDRLDLPAPRPSPSSPNVRYKYLVGAQGRCE----- 93  
 DB 4 GPRYIQAVGP-----IGPRG-PPGPPSPGQO-----GYQGLRGPDSGPMGPIGK 50  
 QY 94 -APPAPMLLTCDPRDLDTFTIKFQYSPNLMGHEFSHHDYIATSDTEGLSISQ 152  
 DB 51 RGPGRP-----AGIAG-----KSGDGDGSGRPR 75  
 QY 153 GGVCLT-----RGMKVL-----LRVGSPRGQ--AVPRKPVSEM 184  
 DB 76 GGIQPMGRGAGWPMGPMGRKGRGFRGLSGKGEQKSGNQPDGPPGAPSGPIGR 135  
 QY 165 PMEDRRAAHSLEBKNNLPDPTSNATSRGAGSPRLPSPMPAVAGAALAL----- 238  
 DB 136 GQTGERG-----RDGKGLPG-----LRGVDLGAPPPPPPISTGSGPPPGPGPSK 183  
 QY 239 -----LLGVAGAGAMCWRRRRAKPSERSHPG-PSGFRGSLGGGGGQ----- 283  
 DB 184 GDRQSGIKAGQIGQGVGSGCGVAGENGHPGMPDMCANEPGASGSGLPSSGSP 243  
 QY 284 GPR-----EAPGELGIALRG-----GAADPFCPHYKVSQDYGHPVYI 324  
 DB 244 GPRMPPTAGSPGQAGAKGCDGPTGEGRPGAPGVKSSGSP-----GDVAGPCHA 294  
 QY 325 VODGPPSP 333  
 DB 295 GRAGKRGSP 303

RESULT 31  
 A54121  
 collagen alpha 4 chain precursor - sea urchin (Strongylocentrotus purpuratus)  
 N/Alternate names: collagen alpha 2(IV) chain homolog  
 C/Species: Strongylocentrotus purpuratus (purple urchin)  
 C/Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 13-Aug-1999  
 C/Accession: A54121; S44317  
 R/Exposito, J.Y.; Suzuki, H.; Georjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.  
 J. Biol. Chem. 269, 13167-13171, 1994  
 A/Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I  
 A/Reference number: A54121; MUID:94250414; PMID:8175744  
 A/Accession: A54121  
 A/Molecule type: mRNA  
 A/Residues: 1-1747 <EXP>  
 C/Cross-references: EMBL:X76730; NID:9483606; PIND:CAA54146.1; PID:9483607  
 C/Genetics:  
 A/Gene: COL4A1pha  
 C/Superfamily: collagen alpha 1(IV) chain

Query Match 5.8%; Score 143; DB 2; Length 1747;  
 Best Local Similarity 24.1%; Pred. No. 0.041;  
 Matches 84; Conservative 24; Mismatches 159; Indels 82; Gaps 14;

QY 7 GPGVRYGALLLGLVGLVNSGLSLPYYNNSAKRFGAEGYLYPQIGDRDLDCPRAR 66

DB 266 GPRGMDG-----MKAGTGVSDLGSYDYGKGLPGYSGERFPGNPGLGLGNMGKGR 321  
 QY 67 PPGPSSPNVRYKYLVGAQGRCEAPAPNLLTCDPRDLDTFTIKFQYSPNLMG 126  
 DB 322 -DGPGRGYYGYK--GPSGYGMDGDPGPPD----- 352  
 QY 127 HEFSHHDYIATSDTEGLSISQVCLTRGMKVLRYVQ-----SPRGAVPRKVS 182  
 DB 353 -----EIVVGVPPEDDPGPNPGRSGALGLFGDQCGYGVPMGPGPPGPTG 405  
 QY 183 EMPMRDRAAHSLEBKNNLPDPTSNATSRGAGSPRLPSPMPAVAGAAL 232  
 DB 406 SQGGRDGEKESGSPGPIPGFQGETGERGNDGNFGERGEKNNMESRPPGMDGSRGRQ 465  
 QY 233 GGLALLLGVAGAGAMCWRRRRAKPSERSHPG-----PSGFRGSLGLGGGGMGP 285  
 DB 466 G-----FMQCKGRGP-----PGRAGPAPASGNSQSFQFGDPTGNTGLKMRGIGALG 516  
 QY 286 REAPGELGIALRGGAADPFCPHYK-VSGDYGHFVYIVODGPPSP 333  
 DB 517 RDGRPSRG-----ELGIC--FPCPPGIKGYPDGRYP-----GDPSP 554

RESULT 32  
 A57131  
 collagen alpha 2(VIII) chain - human (fragment)  
 C/Species: Homo sapiens (man)  
 C/Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 13-Aug-1999  
 C/Accession: A57131  
 R/Mutagaki, Y.; Ucenko, O.; Apte, S.; Mattei, M.G.; Nimmiya, Y.; Olsen, B.R.  
 J. Biol. Chem. 266, 7721-7727, 1991  
 A/Title: The alpha2(VIII) collagen gene. A novel member of the short chain collagen fam  
 A/Reference number: A57131; MUID:91210292; PMID:2019595  
 A/Accession: A57131  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-635 <MUR>  
 A/Cross-references: GB:M60832; NID:9177178; PIND:AAA62822.1; PID:9177179  
 C/Genetics:  
 A/Gene: GDB:COL8A2  
 A/Cross-references: GDB:127812; OMIM:120252  
 A/Map position: 1p34.3-1p32.3  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F/1-11/Domain: amino-terminal nonhelical (fragment) #status predicted <NC>  
 F/12-468/Region: interrupted helical  
 F/469-635/Domain: carboxyl-terminal nonhelical #status predicted <NC>  
 F/508-634/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 5.8%; Score 142.5; DB 2; Length 635;  
 Best Local Similarity 26.1%; Pred. No. 0.015;  
 Matches 74; Conservative 17; Mismatches 85; Indels 107; Gaps 14;

QY 147 GLESLQ-----GYCLTRGMKVLRYVQSPRGAVP-----RKPVSEMERDRGA 193  
 DB 222 GLPQGQPSGAKGPPRRPGGLI-----GPTGYMPGLPQPKGRGAPGAVGLGDRG-- 275  
 QY 194 HSLBKNNLPD-----PTSNAT-----RGAGSPRLPSPMPAVAGA 233  
 DB 276 ---EAGEEDGDEGQPGQGLGPPPLPSAGLPGRRPPGLRGEAGPQGPVPPVIRDDQ 332  
 QY 234 --GLALLLGVAGAGAMCWRRRRAKPSERSHPG-----GSFGRG 272  
 DB 333 PSGLA-----GRGVPG-----ERGLGAGPPTGPTPKGPPGTPRGPGVAGALGQK 382  
 QY 273 GSLG-----GGGGMGR-----EAPGELGIALRG-----GGAD 304  
 DB 363 GDLGPPQPGLRGSPGIPGLQGPAGPITGQGLPGLKGPGLPDPGGRAGRPGTAGPRG 442  
 QY 305 PPGPCHYKVSQDYGHPVYIVODGPPSPPNVYTSISVLEMP 347  
 DB 443 PPGVPSGPIRGPPGP-----GPGAPGAFDETGTAGLHP 479



RESULT 33  
 collagen alpha 1(VIII) chain - mouse  
 A:Accession: S23779  
 C:Species: Mus musculus (house mouse)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: S23779  
 R:Muragaki, Y.; Shioe, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Minomura, Y.  
 Eur. J. Biochem. 207, 895-902, 1992  
 A:Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypeptide  
 A:Reference number: S23779; MUID:92362626; PMID:1495564  
 A:Accession: S23779  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-743 <MOR>  
 A:Cross-references: EMBL:X66976; NID:G50493; PIDN:CAA47387.1; PID:G135953  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:616-742/Domain: complement C1q carboxyl-terminal homology <C1q>

Query Match 5.8%; Score 142.5; DB 1; Length 743;  
 Best Local Similarity 21.3%; Pred. No. 0.017;  
 Matches 101; Conservative 29; Mismatches 136; Indels 209; Gaps 21;

QY 14 GALLIGLVGLVSGLSLEPVYNSANKRFQAEQGV-----LYPQIGRLDLCGRAPPG 69  
 Db 6 GPLQGLGLFIIS-----LNSV--RLIQAGAYYGIKPLPPQIPQIPQIPQIPGLG 55  
 QY 70 ---PH-----SSPYEPYKLY-----L 83  
 Db 56 QQVPMPLGKQGLSVKGMFHXQYKEXEYPLQYKKEIPVPMKKEVVPKKGKKEVPLAS 115  
 QY 84 VQAGAGRCRCE---APPAPNLLITCDPRDLDRFTIKQETSPNMGHEFSHHDYITIA 139  
 Db 116 LRGEGPGRPEPRPGRPP-----PGLPGCMG-----IK 146  
 QY 140 TSDGT-----REGLESIQG--GVCLTRGMKVLIRVQSPRGAVPRKPSSEMERDR 190  
 Db 147 GPRGQGRIGIKRPMGPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMK 196  
 QY 191 G--AAHSL---EPGKENDPGDPTSNATSRGAEGPLPPSM----- 225  
 Db 197 GPRGPRGPRGIGIKRPMGPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMKPMK 255  
 QY 226 -PAVGAAGGLALLLLGVAGAGGAMCMRRRAKPBESHNG----- 265  
 Db 256 PPGMGPGRPVGLPEVGRVGTGPPGQPLGKEPGRPEPRQGLIGVGVQSPGMPGV 315  
 QY 266 --PGSFGSGSLGLGSGG-----MGPRE 287  
 Db 316 GKPGDGGSGRGQGFPGGSGEGGLPGLPRLPGLPGVGRKPGPKHNGIGVPLGR- 374  
 QY 288 AEPGGLIALRGGAADPPPCHEKVSQDGHVYTYQDP-----PQSP 334  
 Db 375 GEKGIAPGPMG-----PPGEGPLPGLIPGMPGPGALGFPKPGEGGVVQGGP 425

## RESULT 34

## CGHUV

collagen alpha 2(V) chain precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Jul-1989 #sequence\_revision 28-Jul-1995 #text\_change 31-Dec-2000  
 C:Accession: A14427; A54555; S43643; A25874; I55239; I59025; A25374; A30017  
 R:Woodbury, D.; Benson-Chanda, V.; Ramirez, F.  
 J. Biol. Chem. 264, 2735-2738, 1989

A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struc  
 A:Reference number: A14427; MUID:89123368; PMID:2914927  
 A:Accession: A14427  
 A:Molecule type: mRNA  
 A:Residues: 1-463 <MOR>  
 A:Cross-references: GB:J04478; NID:G179697; PIDN:AA51859.1; PID:G179698  
 R:Experimental source: placenta  
 R:Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.  
 Gene Expr. 1, 29-39, 1991

A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for  
 A:Reference number: A54555; MUID:92314691; PMID:1820205  
 A:Accession: A54555  
 A:Molecule type: DNA  
 A:Residues: 1-32 <GRE>  
 A:Cross-references: GB:M58529; NID:G180834; PIDN:AA041699.1; PID:G555235  
 R:Noridi-Amell, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Bouillon, M.M.; Be  
 Eur. J. Biochem. 221, 987-995, 1994  
 A:Title: Diversity in the processing events at the N-terminus of type-V collagen.  
 A:Reference number: S43643; MUID:94237164; PMID:8181482  
 A:Accession: S43643  
 A:Molecule type: protein  
 A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>  
 A:Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:G1340175  
 R:Well, D.; Bernard, M.; Gargano, S.; Ramirez, F.  
 Nucleic Acids Res. 15, 181-198, 1987  
 A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibril  
 A:Reference number: A25874; MUID:87146331; PMID:3029669  
 A:Accession: A25874  
 A:Molecule type: mRNA, DNA  
 A:Residues: 398-1496 <WEI>  
 A:Cross-references: GB:M0956; NID:G180427; PIDN:AA52007.1; PID:G180428  
 A:Note: part of this sequence were determined by protein sequencing  
 R:Manuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3399, 1985  
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
 A:Reference number: I59025; MUID:85216505; PMID:3858826  
 A:Accession: I59025  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1002-1226 <R2>  
 A:Cross-references: GB:M0956; NID:G180427; PIDN:AA52007.1; PID:G180428  
 A:Note: part of this sequence were determined by protein sequencing  
 R:Manuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3399, 1985  
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
 A:Reference number: I59025; MUID:85216505; PMID:3858826  
 A:Accession: I59025  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1003-1034 <R2>  
 A:Cross-references: GB:M1135; NID:G179693; PIDN:AA51857.1; PID:G179694  
 A:Note: part of this sequence were determined by protein sequencing  
 R:Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.  
 J. Biol. Chem. 260, 11216-11222, 1985  
 A:Title: Complete primary structure of the human alpha-2 type V procollagen COOH-terminu  
 A:Reference number: A25374; MUID:85289337; PMID:2411731  
 A:Accession: A25374  
 A:Molecule type: mRNA  
 A:Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MYE>  
 A:Cross-references: GB:M1718; NID:G180912; PIDN:AA52058.1; PID:G180913  
 A:Experimental source: normal fibroblasts  
 R:Salpoutras, P.; Schwartz, R.C.; Liddell, A.C.; Salfield, C.S.; Weil, D.; Ramirez, F.  
 Genomics 3, 275-277, 1988  
 A:Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on  
 A:Reference number: A30017; MUID:89138450; PMID:3224983  
 A:Accession: A30017  
 A:Molecule type: DNA  
 A:Residues: 1449-1463, 'E', 1465-1495, 'A', <TSI>  
 A:Cross-references: GB:J03051; NID:G179695; PIDN:AA51558.1; PID:G179696  
 A:Note: the authors translated the codon GAA for residue 1460 as Gln, and GAG for residu  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 are 5-hydroxylated and subsequently O-glycosylated.  
 C:Comment: The amino-terminal propeptide domain appears not to be completely cleaved.  
 C:Genetics:  
 A:Gene: COL5A2  
 A:Cross-references: GB:119064; OMIM:120190  
 A:Map position: 2q31-2q31  
 A:introns: 33/1, 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3  
 C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUV), a  
 alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the  
 length, is formed with desmosine cross-links made from lysine and allysine residues  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer associated with cel





```

Db 257 ----PGHGL-PGPVGLGVGKPGVYTFPPGPQGLGKATGEPKGGPIGVPCVQPPG 311
QY 242 VAGAGAMCWRERRRRAKPESEPHPG-----PSFGGSGS 274
Db 312 IPGIG-----KPGQDGIPEGPGFPGKGEGLPGLGABGLPGIGKPFPGPKGD 361
QY 275 LGGSG-GGGMGPAREAPBELGIALRGGAADP--PFGCHYKVSQDGHPIYIVQDDP-- 329
Db 362 KMGQVFPALGPR-GEKGPISGPGIGSGSPGELPGIF-----GPMGPALISFPFGPKG 414
QY 330 ----PGSPP 334
Db 415 EGGIVPGQGP 425

RESULT 37
S21369
collagen alpha 2 (VI) chain precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C/Accession: S21369; S28808; S13745; S18863
R/IdnaHmi: X.Y.Z.; Bardon, X.Y.Z.; Dani, C
Submitted to the EMBL Data Library, April 1992
A/Reference number: S21369
A/Accession: S21369
A/Molecule type: mRNA
A/Residues: 1-1029 <1BR>
A/Cross-references: EMBL:X65582; NID:g49808; PIDN:CAA6541.1; PID:g49809
R/IdnaHmi: A.; Bertrand, B.; Bardon, S.; Anri, E.Z.; Grimaldi, P.; Allhaud, G.; Dani, C
Biochem. J. 289, 141-147, 1993
A/Title: Cloning of alpha2 chain of type VI collagen and expression during mouse develop
A/Reference number: S28808; MUID:91343659; PMID:8380980
A/Accession: S28808
A/Molecule type: mRNA
A/Residues: 266-1029 <1B2>
A/Cross-references: EMBL:X62332; NID:g49906; PIDN:CAA44206.1; PID:g49907
R/IdnaHmi: C.D.; Jimenez, S.A.
Matrix 11, 1-9, 1991
A/Title: Structure of cDNAs encoding the triple-helical domain of murine alpha-2(VI) col
rate oligonucleotides for generation of novel cDNA clones.
A/Reference number: S13745; MUID:9128374; PMID:11709252
A/Accession: S13745
A/Molecule type: mRNA
A/Residues: 266-267, 'S', 269-294, 'L', 296-600 <CON>
A/Cross-references: GB:I06343; NID:g192671; PID:AAA37441.1; PID:g192672
A/Note: The sequence from Fig. 3 is inconsistent with that from Fig. 4 in having 306-ile
C/Keywords: collagen alpha 1(VI) chain; von Willebrand factor type A repeat homology
C/Keywords: alternative splicing; cartilage; cell binding; connective tissue; disulfide
F/1-28/Domain: signal sequence #status predicted <SIG>
F/29-1029/Product: collagen alpha 2(VI) chain #status predicted <MAT>
F/29-265/Domain: globular #status predicted <NC2>
F/54-222/Domain: von Willebrand factor type A repeat homology <VMA1>
F/266-600/Domain: collagenous #status predicted <COL>
F/376-378/Region: cell attachment (R-G-D) motif
F/436-438/Region: cell attachment (R-G-D) motif
F/498-501/Region: cell attachment (R-G-D) motif
F/508-510/Region: cell attachment (R-G-D) motif
F/549-551/Region: cell attachment (R-G-D) motif
F/601-1029/Domain: cell attachment #status predicted <NC1>
F/623-799/Domain: von Willebrand factor type A repeat homology <VMA2>
F/841-1012/Domain: von Willebrand factor type A repeat homology <VMA3>
F/150,337,640,907,963/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.8%; Score 142; DB 1; Length 1029;
Best Local Similarity 29.7%; Pred. No. 0.027;
Matches 90; Conservative 16; Mismatches 125; Indels 72; Gaps 20;

```

```

Db 319 KGEF-----GSDG-RKAGPLAGNKG---TDGKGKLGRI GPPGCKDPPGSRPDP 364
QY 181 ----VSEMPER-DRGA-AHSLPEKXENLPDPTSNATSRGAEGLPSPMPAVAGAAG 234
Db 365 GYPBAGSGFGEKGDQAKKDSGRFGRGPPDP--GDKSKYQGNNGAPSGPVGXGKGG 423
QY 235 LALLLLGVAGAGAMCWRERRAKP-----SESRHGP-----GSFGGSLGLG 278
Db 424 ----RPPGPKGEFERKDPGTGSGPGSDGFKGEKGDGPPGPRGLAEVSSKGA 475
QY 279 GGQGM-GPREAPBGLGIALRGGAADPFCPHYKVSQDGHPIYIVQDDPSP--PN 335
Db 476 GDRGLGPRGPQ-GALGEGKQSGSRDP-----GDAGPRDSQGPGRKDPGRPG 524
QY 336 IYI 338
Db 525 FSX 527

RESULT 38
S16366
collagen alpha 2 (IV) chain precursor - pig roundworm
C/Species: Ascaris suum (pig roundworm)
C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C/Accession: S16366
R/IdnaHmi: J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A/Title: The complete primary structure of a nematode alpha-2(IV) collagen and the part
A/Reference number: S16366; MUID:91340768; PMID:1714907
A/Accession: S16366
A/Molecule type: mRNA
A/Residues: 1-1763 <JBI>
A/Cross-references: GB:M67507; NID:g159648; PIDN:AAA18014.1; PID:g159649
C/Genetics:
A/Intron: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C/Suprafamily: collagen alpha 1(IV) chain
C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfic
F/1-26/Domain: signal sequence #status predicted <SIG>
F/27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F/27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F/43-1528/Domain: collagenous #status predicted <COL>
F/197-199/Region: cell attachment (R-G-D) motif
F/1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F/1530-1638/Domain: repeat NC1 #status predicted <NC12>
F/1639-1763/Domain: repeat NC1 #status predicted <NC12>
F/31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F/126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F/1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 5.8%; Score 142; DB 2; Length 1763;
Best Local Similarity 25.5%; Pred. No. 0.049;
Matches 97; Conservative 24; Mismatches 127; Indels 132; Gaps 23;

```

QY 259 SESHPGSGRGRGSLGLGCGGMPREAPGELIALRG--GADPPFCPEYKVEG 316  
|||  
Db 309 GRPGLPPLPPGF--KGRGLDGLPLGVPLPGKQKAGPGRGDKKAGRGPPGPF---GQ 362  
QY 317 DYGHPIVIVDDGPPGQSPNPI 336  
|||  
Db 363 EF-----SDGPP-GPPGL 374

RESULT 39  
CGH071  
collagen alpha 1(III) chain precursor - human  
N/Alternate names: procollagen alpha 1(III) chain  
C/Species: Homo sapiens (man)  
C/Date: 24-Apr-1984 #sequence, revision 01-Sep-1995 #ext change 21-Jul-2000  
C/Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90  
R/Prockop, D.J.  
Submitted to the EMBL Data Library, February 1989  
A/Reference number: S05272  
A/Accession: S05272  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-1240, 'V', 1242-1466 <PRC>  
A/Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058  
R/Ala-Koko, L.; Kontusari, S.; Baldwin, C.T.; Kiviniemi, H.; Prockop, D.J.  
Biochem. J. 260, 509-516, 1989  
A/Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of huma  
encees.  
A/Reference number: S04642; MUID:89350838; PMID:2764886  
A/Accession: S04642  
A/Molecule type: mRNA  
A/Residues: 1-1196 <Ala>  
A/Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058  
R/Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.  
Gene 78, 255-265, 1989  
A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene  
A/Reference number: PE0011; MUID:89378752; PMID:2777063  
A/Accession: PE0011  
A/Molecule type: DNA  
A/Residues: 1-176 <SEN>  
A/Cross-references: EMBL:M2639; NID:G180813; PIDN:AAAS2040.1; PID:G180814  
R/Toman, P.D.; Ricci, G.A.; de Crombrughe, B.  
Nucleic Acids Res. 16, 7201, 1988  
A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pro  
A/Reference number: S01726; MUID:88303360; PMID:3405773  
A/Accession: S01726  
A/Molecule type: mRNA  
A/Residues: 1-170 <TOM>  
A/Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061  
A/Note: the authors translated the codon CAG for residue 154 as His  
R/Janezko, R.A.; Ramirez, F.  
Nucleic Acids Res. 17, 6742, 1989  
A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.  
A/Reference number: S04887; MUID:89386015; PMID:2780304  
A/Accession: S04887  
A/Molecule type: mRNA  
A/Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,  
A/Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045  
A/Note: the authors' translation of residues 505-932 is inconsistent with the nucleotide  
R/Seyer, J.M.; Kang, A.H.  
Biochemistry 16, 1158-1164, 1977  
A/Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
A/Reference number: A90399; MUID:77134724; PMID:557335  
A/Accession: A90399  
A/Molecule type: protein  
A/Residues: 'V', 169-225, 228-232, 'P', 234-292, 'D', 294-398 <SEV1>  
A/Experimental source: liver  
A/Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact  
R/Seyer, J.M.  
submitted to the Atlas, December 1977  
A/Reference number: A94562

A/Accession: A94562  
A/Molecule type: protein  
A/Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEV2>  
A/Experimental source: liver  
A/Note: author submitted corrections to A90399  
R/Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldestein, G.; Byers, P.H.  
Am. J. Hum. Genet. 53, 62-70, 1993  
A/Title: Parental somatic and germ-line mosaicism for a multilexon deletion with unusual  
teping.  
A/Reference number: I51868; MUID:93304430; PMID:8317500  
A/Accession: I51868  
A/Status: preliminary; translated from GB/EMBL/DBD  
A/Molecule type: DNA  
A/Residues: 186-194 <MT1>  
A/Cross-references: GB:S62925; NID:G386425; PIDN:AAID1937.1; PID:G4261637  
R/Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.  
Biochem. J. 311, 939-943, 1995  
A/Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL  
A/Reference number: S59511; MUID:96067614; PMID:7487954  
A/Accession: S59511  
A/Molecule type: mRNA  
A/Residues: 302-423 <CH1>  
A/Cross-references: GB:S79877; NID:G1195576; PIDN:AA835615.1; PID:G1195577  
R/Seyer, J.M.; Kang, A.H.  
Biochemistry 17, 3404-3411, 1978  
A/Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr f  
A/Reference number: A90414; MUID:79000343; PMID:687591  
A/Accession: A90414  
A/Molecule type: protein  
A/Residues: 399-675, 'N', 677-727 <SEV3>  
A/Experimental source: liver  
R/lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5256-5259, 1991  
A/Title: G to T transversion at position +5 of a splice donor site causes skipping of t  
A/Reference number: I55349; MUID:91161621; PMID:1672129  
A/Accession: I55349  
A/Status: translated from GB/EMBL/DBD  
A/Molecule type: DNA  
A/Residues: 537-605 <LEB>  
A/Cross-references: GB:M59312; NID:G180815; PIDN:AAAS2041.1; PID:G180816  
R/Seyer, J.M.; Mainardi, C.; Kang, A.H.  
Biochemistry 19, 1583-1589, 1990  
A/Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-C85 from t  
A/Reference number: A90438; MUID:80198282; PMID:6246985  
A/Accession: A90438  
A/Molecule type: protein  
A/Residues: 728-835, 'A', 897-964 <SEV4>  
A/Experimental source: liver  
R/Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Cha  
J. Biol. Chem. 265, 17070-17077, 1990  
A/Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping a  
A/Reference number: A38303; MUID:91009133; PMID:2145268  
A/Accession: A38303  
A/Molecule type: mRNA  
A/Residues: 861-1015 <COL>  
A/Cross-references: GB:U05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AAAS9283.1; PID:  
A/Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos sy  
R/Mankoo, B.S.; Dalglish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A/Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A/Reference number: S02119; MUID:88189827; PMID:3357782  
A/Accession: S02119  
A/Status: translation not shown  
A/Molecule type: mRNA  
A/Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
A/Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054  
R/Seyer, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A/Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-C89 from t  
A/Reference number: A90446; MUID:81208139; PMID:7016180  
A/Accession: A90446  
A/Molecule type: protein  
A/Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157

A:Experimental source: liver  
 R:Idol, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye  
 Nucleic Acids Res. 12, 9383-9394, 1984  
 A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage  
 A:Reference number: A93551, MUID:85087944, PMID:6056827  
 A:Accession: A93551  
 A:Molecule type: mRNA  
 A:Residues: 1065-1155, 'P', 1157-1466 <DOI>  
 A:Cross-references: EMBL:X01655, EMBL:X01742, NID:G29584, PIND:CA25821.1  
 R:Wikulin, M.; Dalglish, R.; Klueve-Seckman, B.; Renard, S.T.; Tolstochey, P.; Brant  
 Biochemistry 25, 1408-1413, 1986  
 A>Title: Human type III collagen gene expression is coordinately modulated with the type  
 A:Reference number: 152393, MUID:86187804, PMID:3754462  
 A:Accession: 152393  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1161-1200 <MIS>  
 A:Cross-references: GB:M13146, NID:G180415, PIND:AAA52003.1, PID:G180416  
 R:Emmanuel, B.S.; Cammizzaro, L.A.; Seyer, J.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
 A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
 A:Reference number: 159025, MUID:85216505, PMID:3856826  
 A:Accession: 179359  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1165-1196 <EMSA>  
 A:Cross-references: GB:M1134, NID:G180417, PIND:AAA52004.1, PID:G180418  
 R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sipola, M.; Ramirez, F.  
 J. Biol. Chem. 260, 4357-4363, 1985  
 A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. F  
 A:Reference number: A92516, MUID:85157600, PMID:2579949  
 A:Accession: A92516  
 A:Molecule type: DNA  
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
 A:Cross-references: GB:M0615, GB:M10793, GB:M10794, GB:M10795, GB:M10796, GB:M10797, GE  
 A:Experimental source: liver  
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr, the codons given f  
 action  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C  
 C:Genetics:  
 A:Gene: GDB:COL3A1  
 A:Cross-references: GDB:118729, OMIM:120180  
 A:Map position: 2q31-2q31  
 A:Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/1, 1337/3, 1418/3  
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
 er of their length, is formed with desmosine cross-links made from lysine and allysine  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer that maintains inte  
 C:Superfamily: collagen alpha 1(III) chain; fibrillar collagen carboxyl-terminal homology  
 C:Keywords: coll; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
 F:1-53/Domain: signal sequence #status predicted <SIG>  
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
 F:154-1221/Produce: collagen alpha 1(III) chain #status predicted <MAT>  
 F:154-157/Region: amino-terminal nonhelical telopeptide  
 F:168-1196/Region: helical  
 F:1091-1093/Region: cell attachment (R-G-D) motif  
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
 F:1222-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F:1284-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:161, 1212/Modified site: allysine (Lys) #status predicted  
 F:263, 264, 860, 977, 1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:263/Binding site: carboxylate (Lys) (covalent) #status experimental  
 F:584, 1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
 F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted

Query Match 5.8%, Score 141.5, DB 1, Length 1466,

Best Local Similarity 27.6%, Pred. No. 0.043;  
 Matches 61; Conservative 13; Mismatches 74; Indels 73; Gaps 10;  
 QY 171 PGGAAPRKRKVSSEMERDRGA---AHSLPEKENTLPDPTSNAT-----SRGAEG 218  
 DB 844 PPGSGPAGPFPQGVKGRGSPGCGAAGFPGARGLPFPDPSNGNPDPGSGSGKDG 903  
 QY 219 PLRP-----PSMPAVAGAAGL-----ALLIGVAGAGMCMWRRA 256  
 DB 904 PPGPAGNTGAPGSPGVSGRKGADGPGEGKSGPAGCPGAPGLGAGTGAAGLAPPG 963  
 QY 257 KPESRRPGP-----GSPRGSGSLGLG--GGNGP-----REAPGELGAL 297  
 DB 964 MPGPSPGPGQGVKSGSGKPGANGSLGGERGPPQGLPGLAGTAGPGRDNGPSSDGLPG 1023  
 QY 298 RCG-----GAADPPCPHEKVSQDGHFVYVQDPP 330  
 DB 1024 RDGSPGKGDRENGSPGAP-----GAPGHP-----GPP 1052  
 RESULT 40  
 T33250  
 hypothetical protein T15B7.3 - Caenorhabditis elegans  
 C/Spectes: Caenorhabditis elegans  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C/Accession: T33250  
 R:Pauley, A.; Gatlung, S.  
 submitted to the EMBL Data Library, September 1997  
 A:Description: The sequence of C. elegans coemid T15B7.  
 A:Reference number: 221135  
 A:Accession: T33250  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-319 <PAU>  
 A:Cross-references: EMBL:AF022985, PIND:AB69959.1, GSDDB:GN00023, CESP:T15B7.3  
 A:Experimental source: strain Bristol N2, clone T15B7  
 C:Genetics:  
 A:Gene: CESP:T15B7.3  
 A:Map position: 5  
 A:Introns: 266/1  
 C:Superfamily: unassigned collagens  
 Query Match 5.8%, Score 141, DB 2, Length 319;  
 Best Local Similarity 25.6%, Pred. No. 0.0067;  
 Matches 90; Conservative 27; Mismatches 111; Indels 124; Gaps 23;  
 QY 13 VGLALLGLVGLGVSGSLT-----EPVWNSANKRFOAGG-----YVLYPQIGDRLDLIC 62  
 DB 1 MSASTLVTVASASGAIYVCFVTGMIFNDINSFYDEKIGLXKPKYEQIA--NQMI 58  
 QY 63 PRAPRPGRPHSSPNEYFYKLYVG-----GAQGRCEA--PAAPNILLTCRPPDDL 111  
 DB 59 PTPRPSSGSS-----FLGRMKRQAECNCGOSRCPPAPPGPP-----GGPG-- 101  
 QY 112 RPTIKQKQSPMLMGHEPFSHSDYTIANSQDTRGLESLOGVCGITGKMYLIRVQGP 171  
 DB 102 ---ARGEALPGIAGQ-----PGGARINPATRGFCIT-----CP 135  
 QY 172 RGVAVPRKRVSEMERDRGAHSLDEPGKXNLPDPTSNATSGAEGPLPPSMVAAGA 231  
 DB 136 AGAPGAPGAP-----PGA-----PGKGNNGQPGAPAGS--GGRGP--PGRGP--AGD 177  
 QY 232 AGGLALLGLVAGAGAMCMRRRAKPSRRPG--RGSRRGG--SLIGGGGGG---MG 284  
 DB 178 AG-----SPQPGHPSPGNPGRGGGSRGTTPGASGRPPGG 214  
 QY 285 PRAPRGEGLALRGGAADP--PFCPHYKVSQDGHFVYVQDPPGSPPPN 335  
 DB 215 PAGA-PGQG---RSGGAGTPEPQGP-----PGPGGPGHSGNDGVPPGPN 257  
 RESULT 41  
 A34246

collagen alpha 1(VIII) chain precursor - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C/Accession: A34246  
 R/Yamaguchi, N.; Benya, P.D.; van der Rest, M.; Nimmla, Y.  
 J. Biol. Chem. 264, 16022-16029, 1989  
 A>Title: The cloning and sequencing of alpha 1(VIII) collagen cDNAs demonstrate that type  
 omains similar to those of type X collagen.  
 A:Reference number: A34246; MIMD:89380159; PMID:2476437  
 A/Accession: A34246  
 A/Molecule type: mRNA  
 A/Residues: 1,744 <YAM>  
 A/Cross-references: GB:J05042; NID:G164895; PIDN:AAA31204.1; PID:G164896  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>  
 F:21-117/Region: amino-terminal nonhelical  
 F:118-571/Region: interrupted helical  
 F:572-744/Region: carboxyl-terminal nonhelical  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 5.8%; Score 141; DB 1; Length 744;  
 Best Local Similarity 23.6%; Pred. No. 0.022;  
 Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;

```

QY 67 PPGHSSPNVEFYLYLVG-----GAQG-RCRCAPAPNLLTCDRPDLRLFT 114
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 198 PPGHGLPG-----IGKRGGLPGQPGAKDRPKPPG----- 233
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 115 IKFOEYSPNLMGHEFRSHHDYIATSDGTR---EGLESLOG--GYCLTRGMKVLRLVG 168
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 234 -----FGLQGPX-----GEKGFQMPGLPKLPGKPPGNGPPGCVGVG 272
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 169 Q-----SPRGAVPRKPVSEMERDRGAHSLPEKENTLPDP--TSN 210
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 273 KPGVTGPPGPGPLGKPGPPEPBPQPTIGVPVQGPPLPGVKGPRQDDGIPQDPGPPG 332
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 211 ATSGAGSPPLPPSPMAVA-----GAAGLALLLGVAGAGAMCWRRRR-----A 256
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 333 KGEQGLPLPGLPGLPGVKGKPPGPKD-----RGIGVGLGALPRGKGPVAGPMG 387
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 257 KPSESRHNG-PGSRFGSGSLG---GGGGMGR-----EAERBELGILRGG--GAAD 304
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 388 PPGSPGLPGLPGVPGALGPPGPKGEGIVGPGGPPGKPGGLGPGKGFLEGVG 447
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 305 PP---FCPEYKVSQDYGH-----PYIVQDGP-----QSPNI 336
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 448 PPGIRGLPGRIPGPKGAGHKGLPGLPGVPLGLPKGPRGI.PGQGLQGPPI 499
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 42  
 S23298  
 collagen alpha 1(VIII) chain - chicken  
 C:Species: Gallus gallus (chicken)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C/Accession: S23298  
 R/Nimmla, Y.; Castegnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; Luvall, P.; Mc  
 in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Pre  
 A>Title: The molecular biology of collagens with short triple-helical domains.  
 A:Reference number: S22243  
 A/Accession: S23298  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1,744 <NN>  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 5.8%; Score 141; DB 1; Length 744;  
 Best Local Similarity 23.6%; Pred. No. 0.022;  
 Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;

QY 67 PPGHSSPNVEFYLYLVG-----GAQG-RCRCAPAPNLLTCDRPDLRLFT 114

```

Db 198 PPGHGLPG-----IGKRGGLPGQPGAKDRPKPPG----- 233
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 115 IKFOEYSPNLMGHEFRSHHDYIATSDGTR---EGLESLOG--GYCLTRGMKVLRLVG 168
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 234 -----FGLQGPX-----GEKGFQMPGLPKLPGKPPGNGPPGCVGVG 272
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 169 Q-----SPRGAVPRKPVSEMERDRGAHSLPEKENTLPDP--TSN 210
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 273 KPGVTGPPGPGPLGKPGPPEPBPQPTIGVPVQGPPLPGVKGPRQDDGIPQDPGPPG 332
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 211 ATSGAGSPPLPPSPMAVA-----GAAGLALLLGVAGAGAMCWRRRR-----A 256
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 333 KGEQGLPLPGLPGLPGVKGKPPGPKD-----RGIGVGLGALPRGKGPVAGPMG 387
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 257 KPSESRHNG-PGSRFGSGSLG---GGGGMGR-----EAERBELGILRGG--GAAD 304
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 388 PPGSPGLPGLPGVPGALGPPGPKGEGIVGPGGPPGKPGGLGPGKGFLEGVG 447
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 305 PP---FCPEYKVSQDYGH-----PYIVQDGP-----QSPNI 336
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 448 PPGIRGLPGRIPGPKGAGHKGLPGLPGVPLGLPKGPRGI.PGQGLQGPPI 499
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 43  
 T30165  
 hypothetical protein F26B1.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
 C/Accession: T30165  
 R/Latrelle, P.; Wamsley, P.; Kramer, J.  
 submitted to the EMBL Data Library, November 1996  
 A:Description: The sequence of C. elegans cosmid F26B1.  
 A:Reference number: Z20748  
 A/Accession: T30165  
 A/Status: preliminary; translated from GB/EMBL/DDBJ  
 A/Molecule type: DNA  
 A/Residues: 1-305 <LAT>  
 A/Cross-references: EMBL:U08444; PIDN:AB37788.1; GSPDB:GN00019; CESP:F26B1.4  
 A:Experimental source: strain Bristol N2; clone F26B1  
 A:Gene: CESP:F26B1.4  
 A:Map position: 1  
 A:Insertions: 51/3; 77/1  
 C:Superfamily: unassigned collagens

Query Match 5.7%; Score 140.5; DB 2; Length 305;  
 Best Local Similarity 25.8%; Pred. No. 0.009;  
 Matches 69; Conservative 22; Mismatches 87; Indels 89; Gaps 15;

```

QY 114 TIKFOEYSPNLMGHEFRSHHDYIATSDG-----TEGLESLOGVCLTRGMKVL 164
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 47 TQHFQETSEKIW-----DDLVPVATGETDRHKHAYLTPAVVPVYVG----- 89
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 165 LRVGSPRGAVPRKPVSEMERDRG-----AHSI-----EPKENTLPDP--TSN 210
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 90 FQTGTANRGNCKRPPGPPGNGDGNRYCULSHNLPPLPDDGEGVFGSSGGDIS 149
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 206 ---DP-----TSNATSGAGSPPLPPSPMAVAGAGLALLLGVAGAGAMCWRRRR 257
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 150 VLHNDPGRACQCPAPMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 198
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 258 PSESRHNG-PGSRFGSGSLGIGGGGM-----GPRAPPELGLALRGGAADPP 307
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 199 PEPGLPVPVDPDFDGPAGPAGNNGVAVRRI.PGPPGPGAG--GDSGV--GGG-----Pg 250
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 308 CPHEKVSQDYGH.PYIVQDGPQSP 334
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 251 SPGLAGAPDGGAP-----GPGGP 270
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 44  
 T22482

hypothetical protein F52B11.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T22482  
 R:Matthews, L.  
 Submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19569  
 A:Accession: T22482  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-304 <WIL>  
 A:Cross-references: EMBL:Z82268; PIDN:CA805195.1; GSPDB:GN00022; CESP:F52B11.4  
 A:Experimental source: clone F52B11  
 C:Genetics:  
 A:Gene: CESP:F52B11.4  
 A:Map position: 4  
 A:Introns: 27/3  
 C:Superfamily: unassigned collagens

Query Match 5.7%; Score 140; DB 2; Length 304;  
 Best Local Similarity 26.8%; Pred. No. 0.0097;  
 Matches 59; Conservative 16; Mismatches 87; Indels 58; Gaps 10;

Qy 142 DGTREGLSLQGVLT-----TRGMKVLRYGSPRGAY-----P 177  
 Db 87 EGVDDSGSAGQGGSCGCLPGAAGPACTPG-----KPRGRPGAAALPQNGRPPAQP 141  
 Qy 178 RKPVEEMW---ERDRGAHSLRP-GKENPGDPTNATSRGAEGLPP-----PSMP 226  
 Db 142 CEPTRPPCKRCPPGAPAPGAPGQDAPGAP-GQSGAGAPGAPGPGKASGAGANP 200  
 Qy 227 AVAAG--GLALLLVGAGAGMCMRRRAKPSRRP-----GPGSPRGSLG 276  
 Db 201 GQAGAPGPGADAGSESIPGAPG-----QAGPGPPGAPGAPGAPGAPGAPG 253  
 Qy 277 LGGGCGMPPREAPRPGIARGGADPPRCPHYEKYSG 316  
 Db 254 PSGAPGPGADNGPAPGAPGQDAGGAKGICPKYCAIDG 293

## RESULT 45

CGHUID

collagen alpha 1(X) chain precursor - human

N:Alternate names: procollagen alpha 1(X) chain

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence\_revision 03-Nov-1995 #text\_change 22-Jun-1999

C:Accession: S26396; S30086; S15826; S18249; A43901; I51870; S21856

R:Reichenberger, E.; Beiler, F.; LuValle, P.; Olsen, B.R.; von der Mark, K.; Bertling, W.

FEBS Lett. 311, 305-310, 1992

A:Title: Genomic organization and full-length cDNA sequence of human collagen X.

A:Reference number: S26396; MUID:93012005; PMID:1138733

A:Accession: S26396

A:Molecule type: DNA

A:Residues: 1-680 &lt;RE1&gt;

A:Cross-references: EMBL:X68952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S4

R:Acte, S.S.

Submitted to the EMBL Data Library, March 1992

A:Reference number: S30085

A:Accession: S30086

A:Molecule type: DNA

A:Residues: 1-680 &lt;APT&gt;

A:Cross-references: EMBL:X65120; NID:923129

A&gt;Note: the initial difference is probably due to translation of an intronic sequence

R:Acte, S.; Mactel, M.G.; Olsen, B.R.

FEBS Lett. 282, 353-396, 1991

A:Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene

A:Reference number: S15826; MUID:91243838; PMID:2037056

A:Accession: S15826

A:Molecule type: DNA

A:Residues: 561-647 'G', 649-666 &lt;AP2&gt;

A:Cross-references: EMBL:X58879; NID:930013; PIDN:CAA1666.1; PID:930014

R:Thomas, J.T.; Creeswell, C.C.; Rash, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant, M.

Biochem. J. 280, 617-623, 1991

A:Title: The human collagen X gene. Complete primary translated sequence and chromosome.  
 A:Reference number: S18245; MUID:92109659; PMID:1764025

A:Accession: S18249

A:Molecule type: DNA

A:Residues: 1-26, 'T', 28-680 &lt;THO&gt;

A:Cross-references: EMBL:X60382; NID:930094; PIDN:CAA4293.1; PID:930095

A&gt;Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-AL

R:Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoes, H.; Bertling, W.

Dev. Biol. 148, 562-572, 1991

A:Title: In situ hybridization studies on the expression of type X collagen in fetal hu

A:Reference number: A43901; MUID:92077285; PMID:1743401

A:Accession: A43901

A:Molecule type: mRNA

A:Residues: 547-656 &lt;RE2&gt;

A:Cross-references: GB:M74050; GB:D57494; NID:9339884; PIDN:AAA61221.1; PID:9553796

A&gt;Note: sequence extracted from NCBI Backbone (NCBI:69012, NCBI:69014)

R:Walls, G.A.; Rash, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.

Am. J. Hum. Genet. 54, 169-178, 1994

A:Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain

pe Schmid.

A:Reference number: I51870; MUID:94136476; PMID:8304336

A:Accession: I51870

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 520-597, 'D', 599-680 &lt;WAL&gt;

A:Cross-references: GB:S68531; NID:9545180; PIDN:AA60615.1; PID:9545181

A&gt;Note: mutant sequence from patient with metaphyseal chondrodysplasia type Schmid

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

ed and subsequently O-glycosylated.

C:Genetics:

A:Gene: GDB:COL10A1

A:Cross-references: GDB:128635; OMIM:120110

A:Map position: 6q21-6q22

A:Introns: 52/1

A&gt;Note: a defect in this gene may cause Schmid metaphyseal chondrodysplasia

C:Complex: type X collagen may be a homotrimer

C:Function:

A:Description: structural component of extracellular fibrous polymer specifically and to

be important for skeletogenesis

C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

C:Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer; hydroxylysine;

F:1-18/Domain: signal sequence #status predicted &lt;SIG&gt;

F:19-880/Product: collagen alpha 1(X) chain #status predicted &lt;MAT&gt;

F:19-56/Domain: amino-terminal nonhelical #status predicted &lt;NC2&gt;

F:57-519/Region: interrupted helical

F:520-680/Domain: amino-terminal nonhelical #status predicted &lt;NC1&gt;

F:553-679/Domain: complement C1q carboxyl-terminal homology &lt;C1Q&gt;

F:617/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 5.7%; Score 140; DB 1; Length 680;

Best Local Similarity 29.1%; Pred. No. 0.024;

Matches 57; Conservative 13; Mismatches 84; Indels 42; Gaps 9;

Qy 171 PRG-----GAVPKP-----VSEWPMERDRGAHSLRPGKXENPG-----DPTGNA 211  
 Db 70 PRGHPPSPSPPKPGYSGPGLGEPGPPGSAVGRGVPGLRPGKXGKSPYPPKDV 129  
 Qy 212 TSGAGELPLPPSPMPAVALAAGLALLLVGAGAGMCMRRRAKPSRRHP-PSGFG 270  
 Db 130 GPAGLPGPRGPPGPGIPGAPG-----ISVPGKPGQGGPFGAGPGRFPGKAPGVPGNMG 186  
 Qy 271 RGGSLGLGGGGMGRREAPRGLIALNG--GAADP-----PCPHYKXSGDVGH 320  
 Db 187 QGEMGYGAPG-----RPGERGLPGPGPTGPGGPGGKRGEMGVGGPGIKXDRGF 239  
 Qy 321 PYTIVDGP--PQSP 334  
 Db 240 PEGMGPIDPPGPGGPP 255

## RESULT 46

B44984





Db 879 -----LBPRAGPAGS-----PGEDEKXGIEGPGQKSGKXGQCGMGFGQDDEAG 925  
 QY 273 -----GSLGLGGGGGNGPREAPFGEGLALRGGAADPPFCPHYKVSQDYGHFYIVQD 327  
 Db 926 PFGPPPIQLQGLPG-PFGEENGVDVPM-----GPPGPP-GPGIRGVPYGEQGLFGAGQD 979  
 QY 328 GP--PQSPENI 336  
 Db 980 GPPGFLGPPGL 990

## RESULT 49

A24450

collagen alpha 2(VIII) chain - bovine (fragment)

N/Alternate names: 50K-A collagen (VIII), Descemet's membrane

C/Species: Bos primigenius taurus (cattle)

C/Date: 31-Mar-1988 #sequence\_revision 05-Apr-1995 #text\_change 03-May-1996

C/Accession: S12898; A24450

R/Mann, K.; Jander, R.; Korsching, E.; Kuehn, K.; Rauterberg, J.

FEBS Lett. 273, 168-172, 1990

A/Title: The primary structure of a triple-helical domain of collagen type VIII from bo

A/Reference number: S12898; MUID:91032171; PMID:2226849

A/Accession: S12898

A/Status: preliminary

A/Molecule type: protein

A/Residues: 1-469 &lt;MAN&gt;

R/Kapoor, R.; Bornstein, P.; Sage, E.H.

Biochemistry 25, 3930-3937, 1986

A/Title: Type VIII collagen from bovine Descemet's membrane: structural characterization

A/Reference number: A90507; MUID:86296625; PMID:3527259

A/Accession: A24450

A/Molecule type: protein

A/Residues: 1-24 &lt;KAP&gt;

C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 5.7%; Score 138.5; DB 2; Length 469;

Best Local Similarity 25.3%; Pred. No. 0.02;

Matches 69; Conservative 17; Mismatches 72; Indels 115; Gaps 14;

QY 147 GHSLSGG-----GVCLTRGMKYLRLVQSGPRGAVP-----RKPYSEMMERDRGAA 193  
 Db 216 GLPGPGPGPGAGGPGRTKPPGLI-----GPTGYGMPGLPGPKDGRGAPVGLGDRG-- 269  
 QY 194 HSLPEKKNLPD-----PTSNAT-----SRGAGPLPPPSMPAVAGAG 233  
 Db 270 ---EPGEDEPGEQPGQLGSPGLPGSAGLPGRGVPQPGXGTPRIGPPVPIRGDQG 326  
 QY 234 ---GLALLLVAGAGAGMCRRRRAKPSRSRPPG-----GSFGRG 272  
 Db 327 PSLGLA---GKPLPG-----ERGLPGAHGPPPTGPGEPFTGRPGGPGAGALGQX 376  
 QY 273 GSLGL-----GGGGMGPR-----EAPGELGIALRG 300  
 Db 377 GDLGLPGQGLNGPGLPGLQGPAGPIGQGLPGLXGEBGLPGPPGKGVGEPGVA----- 432  
 QY 301 GAADPPFCPHYKVSQDYGHFYIVQDPPQSP 333  
 Db 433 GPTGPPGVGSPGLTGPFGP-----GPPGPP 459

## RESULT 50

A41182

collagen alpha 1(II) chain precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 13-Aug-1999

C/Accession: A41182; A44885

R/Metazera, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A/Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A/Accession: A41182

A/Status: preliminary; not compared with conceptual translation

A/Residues: 1-1419 <MET>  
 A/Cross-references: GB:M65161  
 R/Cheah, K.S.; Lau, E.T.; Au, P.X.; Tam, P.P.  
 Development 111, 945-953, 1991  
 A/Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag  
 A/Reference number: A44885; MUID:91347939; PMID:1879363

A/Accession: A44885

A/Molecule type: DNA

A/Residues: 1-28 &lt;CHB&gt;

A/Cross-references: GB:S63190; NID:9234368; PIDN:AA819627.1; PID:9234369

A/Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:P63192)

C/Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology

C/Keywords: alternative splicing; collod coll; extracellular matrix; glycoprotein; trim

P,1191-1419/domain: fibrillar collagen carboxyl-terminal homology &lt;FCC&gt;

Query Match 5.7%; Score 138.5; DB 2; Length 1419;

Best Local Similarity 30.1%; Pred. No. 0.068;

Matches 58; Conservative 13; Mismatches 75; Indels 47; Gaps 9;

QY 173 GGAVPRKPYSEMMERDRGAHSLERKENTL-----PGD--PTSNATSGAGPLPPPS 224  
 Db 430 GGAGPLGPPGSRGAPGNRGP-----PGQGLAGPKAPGERGPGSLAGPKAGANDPGRPE 484  
 QY 225 MPVAGAGGIALLLIGVAGAGAGMCRRRRAKPSRSRPPG-----GSFGRGSLGIGGG 280  
 Db 485 EPLGPGARG-----LGRPDNAGPQKVPBSGAPGEDRPPGPGPGAGGQPGVWGFPP 539  
 QY 281 GGMGPRAEPGEGLIA---LRG-----GGAADPPFCPHYKVSQDYGHFYIVQDGP 329  
 Db 540 KGANGBGRKAGKGLAGAPLRLPLPKDGETGAAAGPPGSPGAGERGEGAP-----GP 593  
 QY 330 -----PQSP 334  
 Db 594 SGFGLPGRPGP 606

Search completed: February 11, 2003, 12:05:56  
 Job time: 46.195 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 11, 2003, 12:05:10 ; Search time 19.491 Seconds  
(without alignments)  
597.393 Million cell updates/sec

Title: US-10-021-121-2

Perfect score: 2450  
Sequence: 1 MGPSPGPGCVGVALLLG.....TTLLKQASVEXAGQHPPL 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

1: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubppa/PCT07\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubppa/US05\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubppa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2450	100.0	455	12 US-10-021-121-2	Sequence 2, Appli
2	1841	75.1	340	12 US-10-021-121-4	Sequence 4, Appli
3	1837	75.0	340	9 US-10-138-787-3	Sequence 3, Appli
4	623	25.4	346	12 US-10-021-121-9	Sequence 9, Appli
5	620.5	25.3	333	10 US-09-754-105-2	Sequence 2, Appli
6	620.5	25.3	333	10 US-09-978-339-2	Sequence 10, Appli
7	620.5	25.3	333	12 US-10-021-121-10	Sequence 10, Appli
8	613.5	25.0	333	9 US-10-138-787-4	Sequence 4, Appli
9	599.5	24.5	345	9 US-10-138-787-5	Sequence 4, Appli
10	489	20.0	89	9 US-10-138-787-13	Sequence 13, Appli
11	489	20.0	89	9 US-09-862-179A-17	Sequence 17, Appli
12	284.5	11.6	92	10 US-09-864-761-48262	Sequence 48262, A
13	196.5	8.0	136	10 US-09-864-761-48257	Sequence 48257, A
14	193.5	7.9	106	10 US-09-925-297-639	Sequence 639, App
15	192	7.8	82	9 US-10-138-787-11	Sequence 11, Appli
16	192	7.8	82	9 US-09-862-179A-15	Sequence 15, Appli
17	191.5	7.8	82	9 US-10-138-787-12	Sequence 12, Appli
18	191.5	7.8	82	10 US-09-862-179A-16	Sequence 16, Appli
19	179	7.3	238	10 US-09-904-954-2	Sequence 2, Appli

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	176	174.5	172.5	169.5	169	168.5	167.5	166	164.5	164.5	164.5	164.5	164.5	164.5	164.5	164.5	160.5	145	141	140.5	130.5	130.5	129	129	129	127	126.5	126.5	126.5	126.5	126.5
	7.2	7.1	7.0	6.9	6.9	6.9	6.8	6.8	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.6	5.9	5.8	5.7	5.3	5.3	5.3	5.3	5.3	5.2	5.2	5.2	5.2	5.2	5.2
	209	233	218	201	201	209	228	205	204	204	204	204	204	204	204	204	228	27	28	638	380	674	595	1806	430	684	403	520	520	520	520
	US-09-921-984-2	US-10-138-787-7	US-09-925-297-510	US-09-904-954-4	US-10-138-787-8	US-10-138-787-6	US-08-578-684-4	US-10-138-787-10	US-10-001-054-18	US-10-028-072-288	US-10-123-904-288	US-10-140-470-288	US-10-175-746-288	US-10-176-918-288	US-10-176-921-288	US-10-176-921-288	US-10-138-787-9	US-09-925-301-1225	US-09-862-179A-2	US-10-001-887-108	US-09-823-240-10	US-09-925-299-979	US-09-854-133-187	US-09-738-973-187	US-09-919-497-56	US-09-823-302-518	US-09-823-240-9	US-09-925-302-689	US-09-978-954-614	US-09-978-697-614	US-09-978-697-614
	Sequence 2, Appli	Sequence 7, Appli	Sequence 510, App	Sequence 4, Appli	Sequence 8, Appli	Sequence 6, Appli	Sequence 4, Appli	Sequence 10, Appli	Sequence 48, Appli	Sequence 288, App	Sequence 288, App	Sequence 288, App	Sequence 288, App	Sequence 288, App	Sequence 288, App	Sequence 288, App	Sequence 9, Appli	Sequence 1225, Ap	Sequence 2, Appli	Sequence 186, App	Sequence 10, Appl	Sequence 979, App	Sequence 187, App	Sequence 187, App	Sequence 56, Appl	Sequence 518, App	Sequence 689, App	Sequence 614, App	Sequence 614, App	Sequence 614, App	

## ALIGNMENTS

RESULT 1  
US-10-021-121-2  
Sequence 2, Application US/10021121  
Patent No. US2002014244A1  
GENERAL INFORMATION:  
APPLICANT: Carae, Ingrid W  
TITLE OF INVENTION: A2-1 Neurotrophic Factor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021.121  
FILING DATE: 06-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/635.130  
FILING DATE: 19-Mar-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-9881  
FAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 455 amino acids



TYPE: Amino Acid  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-021-121-2

Query Match 100.0%; Score 2450; DB 12; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2,5e-171;  
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPMSGPGVGVGALLLLGLVGLVSGLSLEPYMNSANKRFAEGGYLVYPOIGRDL 60  
DB 1 MGPMSGPGVGVGALLLLGLVGLVSGLSLEPYMNSANKRFAEGGYLVYPOIGRDL 60  
QY 61 LCPARPPEGSHSPNTEFYCLYVGAQGRCEAPAPNLLTCDRPDLRLFTIKFOEY 120  
DB 61 LCPARPPEGSHSPNTEFYCLYVGAQGRCEAPAPNLLTCDRPDLRLFTIKFOEY 120  
QY 121 SPNLMGHEFRSHHDYIATSDGTREGLESLOGVCLTRGMKVLRVGSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYIATSDGTREGLESLOGVCLTRGMKVLRVGSPRGAVPRKP 180  
QY 181 VSEMPERDRGAHSLPEKENTPGDPTSNATRGABGLPPSPMPAVAGAAGLALLL 240  
DB 181 VSEMPERDRGAHSLPEKENTPGDPTSNATRGABGLPPSPMPAVAGAAGLALLL 240  
QY 241 GVAAGAGAMCRRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPRAEPGEIGIALRG 300  
DB 241 GVAAGAGAMCRRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPRAEPGEIGIALRG 300  
QY 301 GAADPPCPHYEKVSGDYGHPIYVODGPPQSPNITYSYLWPLIHTIQLFPMRSK 360  
DB 301 GAADPPCPHYEKVSGDYGHPIYVODGPPQSPNITYSYLWPLIHTIQLFPMRSK 360  
QY 361 CSRVTFLFVQYITSTCRMTSFTTLNPSQACRAQMGERRIRKFCWGRILGTALF 420  
DB 361 CSRVTFLFVQYITSTCRMTSFTTLNPSQACRAQMGERRIRKFCWGRILGTALF 420  
QY 421 VLVLLILGLNAGQTLRLORASVEAEGQHGPL 455  
DB 421 VLVLLILGLNAGQTLRLORASVEAEGQHGPL 455

# RESULT 2

US-10-021-121-4  
Sequence 4, Application US/10021121  
Patent No. US2002014244A1

GENERAL INFORMATION:  
APPLICANT: Caras, Ingrid W  
TITLE OF INVENTION: A2-1 Neurotrophic Factor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPacIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021,121  
FILING DATE: 06-Dec-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,130  
FILING DATE: 19-Mar-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1001

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-8674  
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear

SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-021-121-4

Query Match 75.1%; Score 1841; DB 12; Length 340;  
Best Local Similarity 100.0%; Pred. No. 4,3e-127;  
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPMSGPGVGVGALLLLGLVGLVSGLSLEPYMNSANKRFAEGGYLVYPOIGRDL 60  
DB 1 MGPMSGPGVGVGALLLLGLVGLVSGLSLEPYMNSANKRFAEGGYLVYPOIGRDL 60  
QY 61 LCPARPPEGSHSPNTEFYCLYVGAQGRCEAPAPNLLTCDRPDLRLFTIKFOEY 120  
DB 61 LCPARPPEGSHSPNTEFYCLYVGAQGRCEAPAPNLLTCDRPDLRLFTIKFOEY 120  
QY 121 SPNLMGHEFRSHHDYIATSDGTREGLESLOGVCLTRGMKVLRVGSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYIATSDGTREGLESLOGVCLTRGMKVLRVGSPRGAVPRKP 180  
QY 181 VSEMPERDRGAHSLPEKENTPGDPTSNATRGABGLPPSPMPAVAGAAGLALLL 240  
DB 181 VSEMPERDRGAHSLPEKENTPGDPTSNATRGABGLPPSPMPAVAGAAGLALLL 240  
QY 241 GVAAGAGAMCRRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPRAEPGEIGIALRG 300  
DB 241 GVAAGAGAMCRRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPRAEPGEIGIALRG 300  
QY 301 GAADPPCPHYEKVSGDYGHPIYVODGPPQSPNITY 338  
DB 301 GAADPPCPHYEKVSGDYGHPIYVODGPPQSPNITY 338

# RESULT 3

US-10-138-787-3  
Sequence 3, Application US/10138787  
Patent No. US2002017298A1

GENERAL INFORMATION:  
APPLICANT: Holland, Sacha  
APPLICANT: Pawson, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
FILE REFERENCE: 11757,23USMO  
CURRENT APPLICATION NUMBER: US/10/138,787  
CURRENT FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/214,631  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: PCT/CA97/00473  
PRIOR FILING DATE: 1997-07-04  
PRIOR APPLICATION NUMBER: 60/021,272  
PRIOR FILING DATE: 1996-07-05  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3  
LENGTH: 340  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-138-787-3

Query Match 75.0%; Score 1837; DB 9; Length 340;  
Best Local Similarity 99.7%; Pred. No. 8,4e-127;  
Matches 337; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPMSGPGVGVGALLLLGLVGLVSGLSLEPYMNSANKRFAEGGYLVYPOIGRDL 60

Db 1 MGPPSSGGGAVGALLLGLVLSGLSPVYNSANKRFOAGGVLVYPOIGDL 60  
 QY LCPRARPPGPHSSPNYEFYKYLIVGAGRCRCEAPAPNLLTCDRDLRFTIKFOEY 120  
 Db 61 LCPRARPPGPHSSPNYEFYKYLIVGAGRCRCEAPAPNLLTCDRDLRFTIKFOEY 120  
 QY 121 SPNLMGHEFRSHHDYIIATSDGTREGLESLOGGVCLTRGKVLIRVQSPRGAVPRKP 180  
 Db 121 SPNLMGHEFRSHHDYIIATSDGTREGLESLOGGVCLTRGKVLIRVQSPRGAVPRKP 180  
 QY 181 VSEMPERDRGAHSLEPEKENTLPGDPTSNATSRGABEGLPPSPMPAVAGAAGLALLL 240  
 Db 181 VSEMPERDRGAHSLEPEKENTLPGDPTSNATSRGABEGLPPSPMPAVAGAAGLALLL 240  
 QY 241 GVAAGAGACWRRRRRAKPSRHPGSGFRGSGLIGGGGGMGPAREAPGELIALRG 300  
 Db 241 GVAAGAGACWRRRRRAKPSRHPGSGFRGSGLIGGGGGMGPAREAPGELIALRG 300  
 QY 301 GAADPPCPHYEKVSGDYGHVYIVODGPPSPNITY 338  
 Db 301 GAADPPCPHYEKVSGDYGHVYIVODGPPSPNITY 338

# RESULT 4

US-10-021-121-9  
 Sequence 9, Application US/10021121  
 Patent No. US2002014244A1  
 GENERAL INFORMATION:  
 APPLICANT: Carae, Ingrid W  
 TITLE OF INVENTION: A2-1 Neurotrophic Factor  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Genentech, Inc.  
 STREET: 1 DNA Way  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Winpatin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/021,121  
 FILING DATE: 06-Dec-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/635,130  
 FILING DATE: 19-Mar-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Torchia, Ph.D., Timothy E.  
 REGISTRATION NUMBER: 36,700  
 REFERENCE/DOCKET NUMBER: P1001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650/225-8674  
 TELEFAX: 650/952-9881  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 346 amino acids  
 TYPE: Amino Acid  
 TOPOLOGY: Linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
 US-10-021-121-9

Query Match 25.4%; Score 623; DB 12; Length 346;  
 Best Local Similarity 39.2%; Pred. No. 2.5e-38;  
 Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGVVGAALLLGLVLSGL-----SLPEVYNSANKRFOAGGVLVYPOIGDL 61  
 Db 4 PGGRVIGKVLVAVVWALCRLATPLAKWLEPVSNLSLTKFSLGGLVITYKIDKDI 63

QY 62 CPRARPPGPHSSPNYEFYKYLIVGAGRCRCEAPAPNLLTCDRDLRFTIKFOEYS 121  
 Db 64 CPRARAGR-----YKYVLVVRPEQAAAGSTVLDENVLVTCNPEQIGRFTIKFOERS 118  
 QY 122 PNLMGHEFRSHHDYIIATSDGTREGLESLOGGVCLTRGKVLIRVQSPRGAVPRKV 181  
 Db 119 PNLMGHEFRSHHDYIIATSDGTREGLESLOGGVCLTRGKVLIRVQSPRGAVPRKV 181  
 QY 182 SEMPERDRGAHSLEPEKENTLPGDPTSNATSRGABEGLPPSPMPAVAGAAGLALLL 236  
 Db 179 SRSKADVTVMQAQAPSRSLSDSKHETVAQESGP-----GAGGSSGDDP 231  
 QY 237 -----LILLVAGAGCA-----MWRRRRAKPSRHPGSGFRGSGL 277  
 Db 232 GFENSKVALFAAVGACVIFLLIIFLVTLIKLRKRRKHTQO-----RAAASL 282  
 QY 278 ---GGGGMGMPAREAPGELIALRGGAADPPCPHYEKVSGDYGHVYIVODGPPSP 333  
 Db 283 SFLASPKGSGTAGTSPSIIIPLR---TTENNYCPHYEKVSGDYGHVYIVQENPQSP 339  
 QY 334 PNITY 338  
 Db 340 ANITY 344

# RESULT 5

US-09-754-105-2  
 Sequence 2, Application US/09754105  
 Patent No. US20010009768A1  
 GENERAL INFORMATION:  
 APPLICANT: Cerrett, Douglas  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING CYTOKINE DESIGNATED LERK-5  
 FILE REFERENCE: 28232  
 CURRENT APPLICATION NUMBER: US/09/754,105  
 PRIOR FILING DATE: 2001-01-03  
 PRIOR APPLICATION NUMBER: 09/329,531  
 PRIOR FILING DATE: 1999-06-10  
 NUMBER OF SEQ ID NOS: 3  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 2  
 LENGTH: 333  
 TYPE: PRT  
 ORGANISM: homo sapiens  
 US-09-754-105-2

Query Match 25.3%; Score 620.5; DB 10; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 3.7e-38;  
 Matches 133; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

QY 14 GALLLGLVLSGLSPVYNSANKRFOAGGVLVYPOIGDLRDLCPRARPPGPHSS 73  
 Db 14 GYLMVLCRAIKSLVLEPIYNSNSKFLPGQGLVLPQIDKLDICPKV---DSKTV 70  
 QY 74 PNLYEFYKYLIVGAGRCRCEAPAPNLLTCDRDLRFTIKFOEYSPNLMGHEFRSH 133  
 Db 71 GQVEYKVVVMDQADRCTICKENTPLINCAKPPQDIKFTIKFOEYSPNLMGLTFOR 130  
 QY 134 DYTITSDGTREGLESLOGGVCLTRGKVLIRVGO--SPRGAVPRKVSMPER-DR 190  
 Db 131 DYTITSDGTREGLESLOGGVCLTRGKVLIRVGO--SPRGAVPRKVSMPER-DR 190  
 QY 191 GAHSLEPEKENTLPGDPTSNATSRGABEGLPPSPMPAVAGAAGLALLLGVAGAGANC 250  
 Db 191 GAHSLEPEKENTLPGDPTSNATSRGABEGLPPSPMPAVAGAAGLALLLGVAGAGANC 250  
 QY 251 GRSSTSPFVKENPGSGTGNAGSHGNNILGSEVALFAGISGCIIFVITLVVLL 250  
 Db 251 KRRRAKPSRHPGSGFRGSGLIGGGGGMGPAREAPGELIALRGGAADPPCPH 310  
 QY 311 YEKVSGDYGHVYIVODGPPSPNITY 338  
 Db 251 KYRRRRKSPQHTTTLSTLATPRSGNN---NGSEPSDIIIPLR---TADSVCPH 303





US-10-138-787-13

Query Match 20.0%; Score 489; DB 9; Length 89;  
Best Local Similarity 100.0%; Pred. No. 3.1e-29;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 RRRRAKSESRRHPGSGFGRGSLGLGGGGMGPRAEPGELIALRGGAADPPFCPHY 311  
DB 1 RRRRAKSESRRHPGSGFGRGSLGLGGGGMGPRAEPGELIALRGGAADPPFCPHY 60  
QY 312 EKVSQDYGHFVYIVQDGPQSPNNITY 338  
DB 61 EKVSQDYGHFVYIVQDGPQSPNNITY 87

RESULT 11  
US-09-862-179A-17  
Sequence 17, Application US/09862179A  
Patent No. US20020147306A1

GENERAL INFORMATION:  
APPLICANT: Lin, Danny  
APPLICANT: Pawsen, Anthony  
TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS  
TITLE OF INVENTION: AND PRZ DOMAINS  
FILE REFERENCE: MTST-P01-009  
CURRENT APPLICATION NUMBER: US/09/862,179A  
CURRENT FILING DATE: 2001-05-21  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 17  
LENGTH: 89  
TYPE: PRZ  
ORGANISM: Homo sapiens  
US-09-862-179A-17

Query Match 20.0%; Score 489; DB 10; Length 89;  
Best Local Similarity 100.0%; Pred. No. 3.1e-29;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 RRRRAKSESRRHPGSGFGRGSLGLGGGGMGPRAEPGELIALRGGAADPPFCPHY 311  
DB 1 RRRRAKSESRRHPGSGFGRGSLGLGGGGMGPRAEPGELIALRGGAADPPFCPHY 60  
QY 312 EKVSQDYGHFVYIVQDGPQSPNNITY 338  
DB 61 EKVSQDYGHFVYIVQDGPQSPNNITY 87

RESULT 12  
US-09-864-761-48262  
Sequence 48262, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 48262  
LENGTH: 92  
TYPE: PRZ  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL136092.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
OTHER INFORMATION: SWISSPROT HIT: P8172, EVALU 4.00e-51  
OTHER INFORMATION: EST\_HUMAN HIT: BE562822.1, EVALU 3.00e-50  
US-09-864-761-48262

Query Match 11.6%; Score 284.5; DB 10; Length 92;  
Best Local Similarity 54.2%; Pred. No. 2.6e-14;  
Matches 52; Conservative 15; Mismatches 24; Indels 5; Gaps 1;

QY 42 FOAGGYLVLPQIGDRLDLCPRAPPGPHSSPNYERKYLVLGGAGRCAPAPAPNTL 101  
DB 1 FLGSKGLVYPKIDKDIICPRBAARP-----YETKLYLRPEGAACSTVLDENVL 55  
QY 102 LTCRPLDLRFTIKFOEYSPNLMGHEFRSHDYI 137  
DB 56 VTCNRPECEIRFTIKFOEFSNPTWGLEFKCHDYI 91

RESULT 13  
US-09-864-761-48257  
Sequence 48257, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6

```

; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48257
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO A1136092.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EST HUMAN HIT: AA377505.1, EVALUE 2.00e-26
; OTHER INFORMATION: SWISSPROT HIT: P52795, EVALUE 4.00e-40
US-09-864-761-48257

Query Match      8.0%; Score 196.5; DB 10; Length 136;
Best Local Similarity 36.8%; Pred. No. 1.1e-07;
Matches 50; Conservative 13; Mismatches 34; Indels 39; Gaps 5;

QY 230 GAAGGLA-----LLLGAGAGAGA-----MCMRRRAKPSRRHGP 266
DB 11 GASGSSSDPDGFENSKVALPAVAGCVITLLITLTLTKRKRRHRTQ----- 65
QY 267 GSFGRGSLG---GGGGMGPREAREPGLGIALRGGAADPFPCPYKVSQDYGHPV 322
DB 66 ---RAAALSLSTLASPKGSGTAGTETPSDITILPR---TTENNYCPHYKVSQDYGHPV 118
QY 323 YIVDGPPOSPPNITY 338
DB 119 YIVDMPPOSPPNITY 134

RESULT 14
US-09-925-297-639
; Sequence 639, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270

```

```

; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 639
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-639

Query Match      7.9%; Score 193.5; DB 10; Length 106;
Best Local Similarity 42.6%; Pred. No. 1.3e-07;
Matches 46; Conservative 11; Mismatches 26; Indels 25; Gaps 4;

QY 235 LALLLGAVAGAGAGMCMRRRAKPSRRHGPGRGSLG---GGGGMGPREAREP 290
DB 18 LTVLLKLT-----KRRKRHTQ-----RAAALSLSTLASPKGSGTAGTETP 59
QY 291 GELGIALRGGAADPFPCPYKVSQDYGHPYIVQDGPPOSPPNITY 338
DB 60 SDIILPR---TTENNYCPHYKVSQDYGHPYIVQEMPPOSPPNITY 104

RESULT 15
US-10-138-787-11
; Sequence 11, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; FILE REFERENCE: 11757.23USMO
; CURRENT APPLICATION NUMBER: US/10/138,787
; PRIOR FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-11

Query Match      7.8%; Score 192; DB 9; Length 82;
Best Local Similarity 46.2%; Pred. No. 1.2e-07;
Matches 42; Conservative 9; Mismatches 24; Indels 16; Gaps 3;

QY 252 RRRRAKPSRRHGPGRGSLG---GGGGMGPREAREPGLGIALRGGAADPF 307
DB 2 RKRRKRHTQ-----RAAALSLSTLASPKGSGTAGTETPSDITILPR---TTENNY 49
QY 308 CPHYKVSQDYGHPYIVQDGPPOSPPNITY 338
DB 50 CPHYKVSQDYGHPYIVQEMPPOSPPNITY 80

RESULT 16
US-09-862-179A-15
; Sequence 15, Application US/09862179A
; Patent No. US20020147306A1
; GENERAL INFORMATION:
; APPLICANT: Ilin, Danny
; APPLICANT: Pawson, Anthony
; TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
; TITLE OF INVENTION: AND PDZ DOMAINS
; FILE REFERENCE: MTSI-P01-009

```

```
/ CURRENT APPLICATION NUMBER: US/09/862,179A
/ CURRENT FILING DATE: 2001-05-21
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 15
/ LENGTH: 82
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-862-179A-15

Query Match
Best Local Similarity 7.8%; Score 192; DB 10; Length 82;
Pred. No. 1.2e-07;
Matches 42; Conservative 9; Mismatches 24; Indels 16; Gaps 3;

QY 252 RRRRAKPSERHGHGPGSFGGSGI-----GGGGGWRRAHPEGLALRGGAADPPF 307
DB 2 KRRHKKKTQQ-----RAALSLSTLSPKGGSGTGTPEPSDIIIPLR--TTENNY 49
QY 308 CPHYEKVSGDYGHPVYIVQDGPSPSPNIYY 338
DB 50 CPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80

RESULT 17
US-10-138-787-12
/ Sequence 12, Application US/10138787
/ Patent No. US20020172984A1
/ GENERAL INFORMATION:
/ APPLICANT: Holland, Sacha
/ APPLICANT: Mbamalu, Geraldine
/ APPLICANT: Pawsen, Tony
/ TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
/ TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
/ FILE REFERENCE: 11757.23USMO
/ CURRENT APPLICATION NUMBER: US/10/138,787
/ CURRENT FILING DATE: 2002-05-03
/ PRIOR APPLICATION NUMBER: US/09/214,631
/ PRIOR FILING DATE: 1999-03-12
/ PRIOR APPLICATION NUMBER: PCT/CA97/00473
/ PRIOR FILING DATE: 1997-07-04
/ PRIOR APPLICATION NUMBER: 60/021,272
/ PRIOR FILING DATE: 1996-07-05
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 82
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-138-787-12

Query Match
Best Local Similarity 7.8%; Score 191.5; DB 9; Length 82;
Pred. No. 1.4e-07;
Matches 36; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 288 AEPGEGLIALRGGAADPPFCPHYEKVSGDYGHPVYIVQDGPSPSPNIYY 338
DB 33 SEPSDIIIPLR---TADSVFCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80

RESULT 18
US-09-862-179A-16
/ Sequence 16, Application US/09862179A
/ Patent No. US20020147306A1
/ GENERAL INFORMATION:
/ APPLICANT: Lin, Danny
/ APPLICANT: Pawsen, Anthony
/ TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
/ TITLE OF INVENTION: AND PZ DOMAINS
/ FILE REFERENCE: MTSI-P01-009
/ CURRENT APPLICATION NUMBER: US/09/862,179A
/ CURRENT FILING DATE: 2001-05-21
/ NUMBER OF SEQ ID NOS: 44
```

```
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 16
/ LENGTH: 82
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-862-179A-16

Query Match
Best Local Similarity 7.8%; Score 191.5; DB 10; Length 82;
Pred. No. 1.4e-07;
Matches 36; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 288 AEPGEGLIALRGGAADPPFCPHYEKVSGDYGHPVYIVQDGPSPSPNIYY 338
DB 33 SEPSDIIIPLR---TADSVFCPHYEKVSGDYGHPVYIVQEMPPQSPANIYY 80

RESULT 19
US-09-904-954-2
/ Sequence 2, Application US/09904954
/ Patent No. US20020010325A1
/ GENERAL INFORMATION:
/ APPLICANT: BECKMAN, M. P.
/ APPLICANT: CERRETTI, DOUGLAS P.
/ TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
/ RECEPTOR HER
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: IMMUNEX CORPORATION
/ STREET: 51 UNIVERSITY STREET
/ CITY: SEATTLE
/ STATE: WASHINGTON
/ COUNTRY: USA
/ ZIP: 98101
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: Apple Macintosh
/ OPERATING SYSTEM: Apple System 7.1
/ SOFTWARE: Microsoft Word for Apple, Version 5.1a
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/904,954
/ FILING DATE: 12-Jul-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/240,124
/ FILING DATE: <Unknown>
/ APPLICATION NUMBER: US 08/114,426
/ FILING DATE: 30-AUG-1993
/ APPLICATION NUMBER: US 08/109,745
/ FILING DATE: 20-AUG-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: SEESE, KATHRYN A.
/ REGISTRATION NUMBER: 32,172
/ REFERENCE/DOCKET NUMBER: 2814-C
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 587-0430
/ TELEFAX: (206) 233-0644
/ TELEX: 756822
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 238 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-904-954-2

Query Match
Best Local Similarity 7.3%; Score 179; DB 10; Length 238;
Pred. No. 3.8e-06;
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGVVRGALLILGVGLSLPEYVNSAKRFAEGGYLYIQIGRLDILCP--R 64
DB 24 GPG-----GALG-----NRHAYVWNSNOHLRRE-GYTVQVNVVDYLDIYCPHYN 67
```

```

Qy 65 APPGPHSP-----NYEFKYLTVGAGRCCEAPAPNLLITCDRPL---DLRFITKE 117
Db 68 SSGVBPAGAPGGGAEQVLYLWVSRNGRTCNASQGFK-RMECNRPAPHSPIKFSKFK 126
Qy 118 CEYSPNLMGHEFRSHHDYIATSDGTREGLESLOGGVCLTRGMKVLRLVGSQSPRGCAVP 177
Db 127 QRYSAFSLGIEFHAGHEYYIISTPTNHL-----WKCLR--MKVFCVCASTSHSG--- 174
Qy 178 RXPVSEMP-----MERDRGAHSLR-----PGKENLP 204
Db 175 EKPVPFTLPQFTWGPVYKINVLDFEGENPQVPKLEKISGTSFKREHLP 223

RESULT 20
US-09-921-984-2
Sequence 2, Application US/09921984
Patent No. US20020156239A1
GENERAL INFORMATION:
APPLICANT: Pianagan, John G.
TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related
therein
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (txt)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/921,984
FILING DATE: 31-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/308,814
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMT-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-921-984-2

Query Match 7.2%; Score 176; DB 9; Length 209;
Best Local Similarity 29.3%; Pred. No. 5.4e-06;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;
Qy 33 VYVNSANRFOAE-----GGVLYLPQIGDRDLCPRARPPGPHSSPVEFYKYLTVGGA 87
Db 35 VYVNSANRFOVSAVGDGGYTVSRINDYIDICPHGALP-PALRMERYIILYMWGE 93
Qy 88 GGRCEAPAPNLLITCDRPL---DLRFITKEYSBNLWGHFRSHHDYIATSDGT 144
Db 94 GHASCDHRQGRKWEKCNRRPAPAGGPLTFSEKPOLFTFSSIGFEFRGHEYYIATP-- 151
Qy 145 REGLESLOGGVCLTRGMKVLRLVGSQSPRGCAVPKRPVSEMPMERDRGAHSLR 204

```

```

Db 152 ----ENLVDRPCLR--LKYVYR-----PINETLY 174
Qy 205 GDP-----TSNATRGAG 218
Db 175 EAPFIFTSSSCSGLG 192

RESULT 21
US-10-138-787-7
Sequence 7, Application US/10138787
Patent No. US20020172984A1
GENERAL INFORMATION:
APPLICANT: Holland, Sacha
APPLICANT: Mbamalu, Geraldine
APPLICANT: Pawsen, Tony
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
TITLE OF INVENTION: TYROSINE KINASES
FILE REFERENCE: 11757,21USMO
CURRENT APPLICATION NUMBER: US/10/138,787
CURRENT FILING DATE: 2002-05-03
PRIOR APPLICATION NUMBER: US/09/214,631
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: PCT/CA97/00473
PRIOR FILING DATE: 1997-07-04
PRIOR APPLICATION NUMBER: 60/021,272
PRIOR FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 233
TYPE: prp
ORGANISM: Homo sapiens
US-10-138-787-7

Query Match 7.1%; Score 174.5; DB 9; Length 233;
Best Local Similarity 27.4%; Pred. No. 7.8e-06;
Matches 60; Conservative 24; Mismatches 74; Indels 61; Gaps 11;
Qy 20 GVLGVGSLSPVYVNSANKRFQAEQVLYLPQIGDRDLCPR---APPGHSSPN 75
Db 27 GALG-----NRHAVVYVNSNQHRLRE-GYTVQVNVNDYIDICPHYNSGAGPGGGAE 80
Qy 76 YEFYKLYVG-----GAQGR---CEAPAPNLLITCDRPLDLRFITKEYSBNLW 127
Db 81 YVLYVNSRNGYRTCNASQGFKMECNRPAPHSPIKFSKFK 131
Qy 128 EFRSHHDYIATSDGTREGLESLOGGVCLTRGMKVLRLVGSQSPRGCAVPKRPVSEMP-- 185
Db 132 EFRHAGHEYYIISTPTNHL-----WKCLR--MKVFCVCASTSHSG---EKPVPFTLPQF 179
Qy 186 -----MERDRGAHSLR-----PGKENLP 204
Db 180 TWGPNVYKINVLDFEGENPQVPKLEKISGTSFKREHLP 218

RESULT 22
US-09-925-297-510
Sequence 510, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 510

```



LENGTH: 218  
 TYPE: PR1  
 ORGANISM: Homo sapiens  
 US-09-925-297-510

Query Match 7.0%; Score 172.5; DB 10; Length 218;  
 Best Local Similarity 28.3%; Pred. No. 1e-05;  
 Matches 51; Conservative 30; Mismatches 80; Indels 19; Gaps 6;

QY 8 PGGRVGLALLGLVGLSL---EPVYNSANKRFOAGGYLYPQIGRLDLCPR 64  
 DB 8 PGGRVGLALLGLVGLSL---EPVYNSANKRFOAGGYLYPQIGRLDLCPR 66  
 QY 65 ARPPGHSSBN---YFYKYLIVGAGRCRCEAPAPNLLITCDRDL---DIRTFIKQE 119  
 DB 67 YE---DHSVADAMEQYIIYLVEHEEYQLCQPOSXQVWQCNCRPSAKHPEKSEKFR 123  
 QY 120 YSPNMGHEFRSHDYIATSDGTREGLESLOGGYCLTRGMKVLRLVQSPRGGAVERK 179  
 DB 124 FTFTLTKRPEKSHSYIISKPIHQHEDR-----CLRLKVTYSGKTIHSPQAHDPDE 176

## RESULT 23

US-09-904-954-4  
 Sequence 4, Application US/09904954  
 Patent No. US20020010325A1

GENERAL INFORMATION:  
 APPLICANT: BECKMANN, M. P.  
 CERESETTI, DOUGLAS P.

TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
 RECEPTOR HEK

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: IMMUNEX CORPORATION  
 STREET: 51 UNIVERSITY STREET

CITY: SEATTLE  
 STATE: WASHINGTON  
 COUNTRY: USA

ZIP: 98101

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Apple System 7.1

SOFTWARE: Microsoft Word for Apple, Version 5.1a

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/904,954  
 FILING DATE: 12-Jul-2001

CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/240,124  
 FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/114,426  
 FILING DATE: 30-AUG-1993

APPLICATION NUMBER: US 08/109,745  
 FILING DATE: 20-AUG-1993

ATTORNEY/AGENT INFORMATION:  
 NAME: SERSE, KATHRYN A.  
 REGISTRATION NUMBER: 32,172

REFERENCE/DOCKET NUMBER: 2814-C  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644

INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 201 amino acids  
 TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: Protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-904-954-4

Query Match 6.9%; Score 169.5; DB 10; Length 201;

Best Local Similarity 29.9%; Pred. No. 1.5e-05;  
 Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVGSLG---EPVYNSANKRFOAGGYLYPQIGRLDLCPRARPPGHSSPYEYKY 82  
 DB 20 LRGGSLRHVYVNSNPR-LRGDAVELGLNDYLDIVCPHYEGP---ETPALT 76  
 QY 83 LVGAGRCRCEAP-PAPNLLITCDRDLRTFTIKFOEYSPNMGHEFRSHDYIAT 141  
 DB 77 MVDWPGYESQAGSPRAYKWCISLPGHVFSEKIQRTFSLGFEFLGETTYISVP 136  
 QY 142 DGTREGLESLOGGYCLTRGMKVLRLVQSPRGGAVERKPYSEMPERDGAASLEPGE 201  
 DB 137 --TPE-----SSGQCL-----RLQVSCCKRKSAPV----- 164  
 QY 202 NLPGDPTSNATS--RGAEGPLPPSPMPAVAGAALALLL 240  
 DB 165 ---GSPGESGTSGMRGDTSP-----LCILL 189

## RESULT 24

US-10-138-787-8  
 Sequence 8, Application US/10138787  
 Patent No. US20020172984A1

GENERAL INFORMATION:  
 APPLICANT: Holland, Sacha

APPLICANT: Madala, Geraldine

APPLICANT: Pawsan, Tony  
 TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
 BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

TITLE OF INVENTION: TYROSINE KINASES

FILE REFERENCE: 11757.23USWO

CURRENT APPLICATION NUMBER: US/10/138,787

CURRENT FILING DATE: 2002-05-03

PRIOR APPLICATION NUMBER: US/09/214,631

PRIOR FILING DATE: 1999-03-12

PRIOR APPLICATION NUMBER: PCT/CA97/00473

PRIOR FILING DATE: 1997-07-04

PRIOR FILING DATE: 1996-07-05

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 201

TYPE: PR1

ORGANISM: Homo sapiens

US-10-138-787-8

Query Match 6.9%; Score 169; DB 9; Length 201;  
 Best Local Similarity 29.8%; Pred. No. 1.7e-05;  
 Matches 70; Conservative 19; Mismatches 84; Indels 62; Gaps 11;

QY 17 LILGVL-----GLVGLSL-EPVYNSANKRFOAGGYLYPQIGRLDLCPRAPP 68  
 DB 6 LILTVMAAFGLGSPRLGSSLRHVYVNSNPR-SLRGDAVELGLNDYLDIVCPHYEG 64

QY 69 GPPSSPYEYKYLYVGAQRCRCEAP-PAPNLLITCDRDLRTFTIKFOEYSPNMGH 127  
 DB 65 GPPEGP---ETPALTMDWPGYESQAGSPRAYKWCISLPGHVFSEKIQRTFSLGFE 122

QY 128 EFRSHDYIATSDGTREGLESLOGGYCLTRGMKVLRLVQSPRGGAVERKPYSEMP 187  
 DB 123 EFLPGETTYISVP---TPE-----SSGQCL-----RLQVSCCKRKSAPV 155

QY 188 RDGAASLEPGEKNI-PGDPSTNATS--RGAEGPLPPSPMPAVAGAALALLL 240  
 DB 156 RKESAPV-----GSPGESGTSGMRGDTSP-----LCILL 189

## RESULT 25

US-10-138-787-6  
 Sequence 6, Application US/10138787  
 Patent No. US20020172984A1

GENERAL INFORMATION:  
APPLICANT: Holland, Sacha  
APPLICANT: Mbamalu, Geraldine  
APPLICANT: Pawson, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
TITLE OF INVENTION: TYROSINE KINASES  
FILE REFERENCE: 11757.23USMO  
CURRENT APPLICATION NUMBER: US/10/138,787  
CURRENT FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/214,631  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: PCT/CA97/00473  
PRIOR FILING DATE: 1997-07-04  
PRIOR APPLICATION NUMBER: 60/021,272  
PRIOR FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 209  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-138-787-6

Query Match 6.9%; Score 168.5; DB 9; Length 209;  
Best Local Similarity 35.9%; Pred. No. 1.9e-05;  
Matches 42; Conservative 14; Mismatches 52; Indels 9; Gaps 3;

QY 33 VYVNSANKRFOAE-----GGYLYVQIGDRDLDCPRAPPGPHSSPNYERYKYLVGA 87  
DB 35 VYVNSANKRFOAE-----GGYLYVQIGDRDLDCPRAPPGPHSSPNYERYKYLVGA 93  
QY 88 GGRRCCEAPPAPNLLTCDRPLD--DLRFTIKFOEYSPNLMGHEFRSHDDYIAT 141  
DB 94 GNASCDHRGRGRKRCNRPAPGGLPKFSEKQFLTFPSLGFEPPGHEYYIAT 150

RESULT 26  
US-08-578-684-4  
Sequence 4, Application US/08578684  
Patent No. US20020137126A1  
GENERAL INFORMATION:  
APPLICANT: Caras, Ingrid W.  
APPLICANT: Wainslow, John W.  
TITLE OF INVENTION: AL-1 Neurotrophic Factor  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/578,684  
FILING DATE: 02-Jan-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/330128  
FILING DATE: 27-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/486449  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P0920P2  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-8674  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 228 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-578-684-4

Query Match 6.8%; Score 167.5; DB 8; Length 228;  
Best Local Similarity 28.8%; Pred. No. 2.5e-05;  
Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;

QY 33 VYVNSANKRFOAGGYLYVQIGDRDLDCPRAPPGPHSSPNYERYKYLVG----- 85  
DB 34 VYVNSANKRFOAGGYLYVQIGDRDLDCPRAPPGPHSSPNYERYKYLVG----- 90  
QY 86 -----GAQRRCCEAPPAPNLLTCDRPLD--DLRFTIKFOEYSPNLMGHEFRSHDDYIAT 140  
DB 91 DHTSKFGRKRCNRPSPN-----GPKFSEKQFLTFPSLGFEPPGHEYYIAT 141  
QY 141 S-----DTRGLESLOGGVCLTRGKYLRLVGGSPRGANVRKVSMPMRDRGAHSLE 197  
DB 142 AIPDNGRRS-----CLK--LKVFR---PTNSCKKIGVHVRVDPVDKVENSIDE 186  
QY 198 PKENLPGDPTSNATSGAEGPLPPSPMPAVAGAAGLALLLGA 243  
DB 187 PADDIV---HESAPSRG--ENAAQTRIPSRLL---LAILFLLA 223

RESULT 27  
US-10-138-787-10  
Sequence 10, Application US/10138787  
Patent No. US20020172984A1  
GENERAL INFORMATION:  
APPLICANT: Holland, Sacha  
APPLICANT: Mbamalu, Geraldine  
APPLICANT: Pawson, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
TITLE OF INVENTION: TYROSINE KINASES  
FILE REFERENCE: 11757.23USMO  
CURRENT APPLICATION NUMBER: US/10/138,787  
CURRENT FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/214,631  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: PCT/CA97/00473  
PRIOR FILING DATE: 1997-07-04  
PRIOR APPLICATION NUMBER: 60/021,272  
PRIOR FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10  
LENGTH: 205  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-138-787-10

Query Match 6.8%; Score 166; DB 9; Length 205;  
Best Local Similarity 27.5%; Pred. No. 2.8e-05;  
Matches 46; Conservative 31; Mismatches 74; Indels 16; Gaps 5;

QY 18 LIGVGLVSGLSLEPYVNSANKRFOAGGYLYVQIGDRDLDCPRAPPGPHSSPN-- 75  
DB 8 LIGLCCSLAARHRTVFNNSNPKFENE--DYTIHQVLDYDIICPHYE--DHSVADA 63  
QY 76 YERYKYLVGAQGRRCCEAPPAPNLLTCDRPLD--DLRFTIKFOEYSPNLMGHEFRSH 132  
DB 64 MEQYIILYVHEHEYQICQSQKDQVAMQCNRPASAKHGPKLSEKQRTPTFLGSEFKEG 123  
QY 133 HDYIATSDGTRGLESLOGGVCLTRGKYLRLVGGSPRGANVRK 179  
DB 124 HSYVYISKPIHGHDR-----CLRLKTVSGKITHSPQAVNPOE 163



```

; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30939
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00376
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04341
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/06894
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/22031
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/30873
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06666
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/17092
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: PCT/US01/27039
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 48
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-001-054-48

Query Match      6.7%  Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

Cy 18 LIGVLGVSGLSLEFVYNSANKRQAGGYVLPQIDRLDLCPRAPRPHSS--PNY 76
Db 8 LIGLCCSLAADRHVTFVFNSSNPKFRNE-DYTIHVQLNDYDIDICPHYE--DHSADAM 63
Cy 77 EYFKLYLVGAGQRCGEAPRPNLLITGDRPDL--DLRFITKQESPNLMGHEFRSH 133
Db 64 EYIILVHEHEVYQLCQOSQSDQVWQCNRPSSAKGPEKLSKFKRTPTLTGHEFVEGH 123
Cy 134 DYLIIATSDGTREGLSESLQGVCLTRGKKVLLRVGQSPRGGAVERK 179
Db 124 SYVYSKPIHNEHR-----CLFKYTVSGKITHSPQADNPQ 162
```

```

RESULT 29
US-10-028-072-288
; Sequence 288, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; PRIOR APPLICATION NUMBER: 2001-12-19
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704
```

;; PRIOR FILING DATE: 1997-10-29  
;; PRIOR APPLICATION NUMBER: 60/063733  
;; PRIOR FILING DATE: 1997-10-29  
;; PRIOR APPLICATION NUMBER: 60/063735  
;; PRIOR FILING DATE: 1997-10-29  
;; PRIOR APPLICATION NUMBER: 60/063738  
;; PRIOR FILING DATE: 1997-10-29  
;; PRIOR APPLICATION NUMBER: 60/063755  
;; PRIOR FILING DATE: 1997-10-17  
;; PRIOR APPLICATION NUMBER: 60/064248  
;; PRIOR FILING DATE: 1997-11-03  
;; PRIOR APPLICATION NUMBER: 60/064809  
;; PRIOR FILING DATE: 1997-11-07  
;; PRIOR APPLICATION NUMBER: 60/065186  
;; PRIOR FILING DATE: 1997-11-12  
;; PRIOR APPLICATION NUMBER: 60/065846  
;; PRIOR FILING DATE: 1997-11-17  
;; PRIOR APPLICATION NUMBER: 60/066364  
;; PRIOR FILING DATE: 1997-11-21  
;; PRIOR APPLICATION NUMBER: 60/066453  
;; PRIOR FILING DATE: 1997-11-24  
;; PRIOR APPLICATION NUMBER: 60/066511  
;; PRIOR FILING DATE: 1997-11-24  
;; PRIOR APPLICATION NUMBER: 60/066770  
;; PRIOR FILING DATE: 1997-11-24  
;; PRIOR APPLICATION NUMBER: 60/069212  
;; PRIOR FILING DATE: 1997-12-11  
;; PRIOR APPLICATION NUMBER: 60/069278  
;; PRIOR FILING DATE: 1997-12-11  
;; PRIOR APPLICATION NUMBER: 60/069334  
;; PRIOR FILING DATE: 1997-12-11  
;; PRIOR APPLICATION NUMBER: 60/069694  
;; PRIOR FILING DATE: 1997-12-16  
;; PRIOR APPLICATION NUMBER: 60/072320  
;; PRIOR FILING DATE: 1998-01-23  
;; PRIOR APPLICATION NUMBER: 60/073612  
;; PRIOR FILING DATE: 1998-02-04  
;; PRIOR APPLICATION NUMBER: 60/074086  
;; PRIOR FILING DATE: 1998-02-09  
;; PRIOR APPLICATION NUMBER: 60/074092  
;; PRIOR FILING DATE: 1998-02-09  
;; PRIOR APPLICATION NUMBER: 60/07791  
;; PRIOR FILING DATE: 1998-03-12  
;; PRIOR APPLICATION NUMBER: 60/078910  
;; PRIOR FILING DATE: 1998-03-20  
;; PRIOR APPLICATION NUMBER: 60/079294  
;; PRIOR FILING DATE: 1998-03-25  
;; PRIOR APPLICATION NUMBER: 60/079663  
;; PRIOR FILING DATE: 1998-02-27  
;; PRIOR APPLICATION NUMBER: 60/079728  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/080165  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/081203  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081229  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081695  
;; PRIOR FILING DATE: 1998-04-14  
;; PRIOR APPLICATION NUMBER: 60/081817  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081818  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082999  
;; PRIOR FILING DATE: 1998-04-24  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083545  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/084600  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084627  
;; PRIOR FILING DATE: 1998-05-07

;; PRIOR APPLICATION NUMBER: 60/084637  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085149  
;; PRIOR FILING DATE: 1998-05-12  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/086414  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/086430  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087106  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/088026  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088730  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088741  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088810  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088858  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089532  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089599  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089907  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089947  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/090349  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090429  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090538  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07

Query Match 6.7%; Score 164.5; DB 9; Length 204;  
Best Local Similarity 27.7%; Pred. No. 3.6e-05;  
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

OY 18 LITGLVIGSLSPVYVNSANKRFQAEAGVLYPQIGRLDLCPPARPPGPHSS-PNY 76  
DB 8 LITGCCSLAADRHTVTVNSNPKFRNE-DYTHVQUNDYVDIICPAYE--DHSADAM 63  
OY 77 EFKYLVGAGQRCCEAPFAPNLLTCDBPDL--DLRFIKFOEYSPNLMGHEFRSH 133  
DB 64 EGYILVYHEHYQLCQPOSMDYRQCNRPSSAKHGPEKLSKFRQRTPTLTIGHEFRGH 123  
OY 134 DYIITATSDGTRREGLSLGGVCLTRGMKYLKRVGSPRGANVPRK 179  
DB 124 SYIYSKPIHQBDR-----CLRMLVTYSGKITHSFOAHNDPOE 162

```
RESULT 30
US-10-121-049-288
; Sequence 288, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; PRIOR APPLICATION: 2002-04-12
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-288

Query Match      6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LIGVLGVSGLSLEPYVNSANKRFOAEGGYLVYPOIGRDLDCPRARPPGPHSS-PNY 76
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 8 LIGLCCSLAABDHTVFNSSNPKFRNE-DYTIHVQLNDYDIICPHYE--DHSADAM 63
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
QY 77 EFKYLYVGAQGRCEAPPAENLLTCORPDL--DLRFTIKFOEYSPNLMGHEFRSH 133
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 64 EGYTLVVEHEEYQLCPQSKQVRCNRPAAKHPKSEKFORFTPTLKEFEKGGH 123
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
QY 134 DYIITATSDGTREGLESLOGGVCLTRGMKVLAVGSPRGANPRK 179
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 124 SYIYSKPIHQHEDR-----CLRKLKTVSGKITHSPOAHNDPQE 162
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:

RESULT 31
US-10-123-904-288
; Sequence 288, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
```

```
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; PRIOR APPLICATION: 2002-04-16
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-904-288

Query Match      6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LIGVLGVSGLSLEPYVNSANKRFOAEGGYLVYPOIGRDLDCPRARPPGPHSS-PNY 76
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 8 LIGLCCSLAABDHTVFNSSNPKFRNE-DYTIHVQLNDYDIICPHYE--DHSADAM 63
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
QY 77 EFKYLYVGAQGRCEAPPAENLLTCORPDL--DLRFTIKFOEYSPNLMGHEFRSH 133
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 64 EGYTLVVEHEEYQLCPQSKQVRCNRPAAKHPKSEKFORFTPTLKEFEKGGH 123
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
QY 134 DYIITATSDGTREGLESLOGGVCLTRGMKVLAVGSPRGANPRK 179
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 124 SYIYSKPIHQHEDR-----CLRKLKTVSGKITHSPOAHNDPQE 162
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:

RESULT 32
US-10-140-470-288
; Sequence 288, Application US/10140470
; Publication No. US2003002231A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C16
; CURRENT APPLICATION NUMBER: US/10/140,470
; PRIOR APPLICATION: 2002-05-06
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-288

Query Match      6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LIGVLGVSGLSLEPYVNSANKRFOAEGGYLVYPOIGRDLDCPRARPPGPHSS-PNY 76
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 8 LIGLCCSLAABDHTVFNSSNPKFRNE-DYTIHVQLNDYDIICPHYE--DHSADAM 63
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
QY 77 EFKYLYVGAQGRCEAPPAENLLTCORPDL--DLRFTIKFOEYSPNLMGHEFRSH 133
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 64 EGYTLVVEHEEYQLCPQSKQVRCNRPAAKHPKSEKFORFTPTLKEFEKGGH 123
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
QY 134 DYIITATSDGTREGLESLOGGVCLTRGMKVLAVGSPRGANPRK 179
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
DB 124 SYIYSKPIHQHEDR-----CLRKLKTVSGKITHSPOAHNDPQE 162
   |||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:  ::|||:
```

```
Db 8 LLGLCSLAADRHVFNSSNPKFNE-DYTIHVQNDYDILCPHYE---DHSADAM 63
QY 77 EFYLYLVGAGRCAPPAAPNLLTCDRPD---DLRTIKQESPLMGEHFRSH 133
Db 64 EQLYLYVHEHEYOQCPQSDQVRWCNRPASAGPEKSEKFORPTTLGKEFEHGH 123
QY 134 DYIYIATSDGTREGLESIGQGVCTLRGMKYLRLVQSPRGAVPRK 179
Db 124 SYIYISKPIHQHEDR-----CLRKLVTYSGKITTHSPQAHNDPQE 162

RESULT 33
US-10-175-746-288
; Sequence 288, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Geo, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C353
; CURRENT APPLICATION NUMBER: US/10/175,746
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-288

Query Match 6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LLGLVGLVSGLSLEPVYVNSANKRFOAGGYVLYPQIGRLDLCPAPRPPGPHSS-PNY 76
Db 8 LLGLCSLAADRHVFNSSNPKFNE-DYTIHVQNDYDILCPHYE---DHSADAM 63
QY 77 EFYLYLVGAGRCAPPAAPNLLTCDRPD---DLRTIKQESPLMGEHFRSH 133
Db 64 EQLYLYVHEHEYOQCPQSDQVRWCNRPASAGPEKSEKFORPTTLGKEFEHGH 123
QY 134 DYIYIATSDGTREGLESIGQGVCTLRGMKYLRLVQSPRGAVPRK 179
Db 124 SYIYISKPIHQHEDR-----CLRKLVTYSGKITTHSPQAHNDPQE 162

RESULT 34
US-10-176-918-288
; Sequence 288, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
```

```
APPLICANT: Geo, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-288

Query Match 6.7%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 3.6e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LLGLVGLVSGLSLEPVYVNSANKRFOAGGYVLYPQIGRLDLCPAPRPPGPHSS-PNY 76
Db 8 LLGLCSLAADRHVFNSSNPKFNE-DYTIHVQNDYDILCPHYE---DHSADAM 63
QY 77 EFYLYLVGAGRCAPPAAPNLLTCDRPD---DLRTIKQESPLMGEHFRSH 133
Db 64 EQLYLYVHEHEYOQCPQSDQVRWCNRPASAGPEKSEKFORPTTLGKEFEHGH 123
QY 134 DYIYIATSDGTREGLESIGQGVCTLRGMKYLRLVQSPRGAVPRK 179
Db 124 SYIYISKPIHQHEDR-----CLRKLVTYSGKITTHSPQAHNDPQE 162

RESULT 35
US-10-176-921-288
; Sequence 288, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Geo, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 288
; LENGTH: 204
```

;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-176-921-288

Query Match 6.7%; Score 164.5; DB 9; Length 204;  
Best Local Similarity 27.7%; Pred. No. 3.6e-05;  
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LIGVLGLVGLLEPYWMSANKRFOAGEGYLYVPOIDRLDLCPRARPPGPHSS-PNY 76  
DB 8 LIGLCCSLAADRRTVFMSSNPKFRNE-DYTHVQINDYDIICPHYE---DSADAM 63  
QY 77 EFKYLYVGAQRCEAPAPAPNLLITCDRPLD---DLFTIKFQESYSPNMGHFRSH 133  
DB 64 EGYIILYVHEHYQCCQPSKQVAMQCNRPASAKHPEKLSKFPQRTFPLGHEFKGH 123  
QY 134 DYIITATSDGTREGLESLOGGYCLTRGKVTLLRVGQSPRGGAVPK 179  
DB 124 SYIYSKPIHQHEDR-----CLRLKTVSGKLTHTSPQADNDPQE 162

RESULT 36  
US-10-138-787-9  
; Sequence 9, Application US/10138787  
; Patent No. US20020172984A1  
; GENERAL INFORMATION:  
; APPLICANT: Holland, Sacha  
; APPLICANT: Mbamalu, Geraldine  
; APPLICANT: Pawsen, Tony  
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
; FILE REFERENCE: 11757.23USWO  
; CURRENT APPLICATION NUMBER: US/10/138,787  
; PRIOR FILING DATE: 2002-05-03  
; PRIOR APPLICATION NUMBER: US/09/214,631  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: PCT/CA97/00473  
; PRIOR FILING DATE: 1997-07-04  
; PRIOR APPLICATION NUMBER: 60/021,272  
; PRIOR FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-138-787-9

Query Match 6.6%; Score 160.5; DB 9; Length 228;  
Best Local Similarity 28.3%; Pred. No. 8e-05;  
Matches 64; Conservative 29; Mismatches 82; Indels 51; Gaps 11;

QY 33 YVMSANRQFAEGGYLYVPOIGRLDLCPRARPPGPHSSPNTEFKYLYV----- 85  
DB 34 YVMSNRRFQ-RGDYHADVCLINDYDFCHYEDSVPEDET--ERYVLHYVDFGYSAC 90  
QY 86 -----GAQGRCEAPAPAPNLLITCDRPLDIFRTIKFQESYSPNMGHFRSHDYIAT 140  
DB 91 DHTSKGFKEWECNRPSPN-----GPKFSEKFLFPFELGFEFPFGREYFISS 141  
QY 141 S---DGTREGLESLOGGYCLTRGKVTLLRVGQSPRGGAVPKPSNEMERDGAHSL 197  
DB 142 AIPNGRRS-----CLK--LKVFR---FTNSCMKTIQVHDVDFVNDKVENSL 186  
QY 198 PEKENLPDPTSNATSRGAEGLPPPSNPAVAGAGLALLLGLVA 243  
DB 187 PADDTV---HESAEPSRG-ENNAQTPRIPSRLL-----LAILFLIA 223

RESULT 37  
US-09-925-301-1225  
; Sequence 1225, Application US/09925301

; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1225  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1225

Query Match 5.9%; Score 145; DB 10; Length 27;  
Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 VSGDYGHPIYIVQDGPQSPPNIIY 338  
DB 1 VSGDYGHPIYIVQDGPQSPPNIIY 25

RESULT 38  
US-09-862-179A-2  
; Sequence 2, Application US/09862179A  
; Patent No. US20020147306A1  
; GENERAL INFORMATION:  
; APPLICANT: Lin, Danny  
; APPLICANT: Pawsen, Anthony  
; TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS  
; TITLE OF INVENTION: AND PDZ DOMAINS  
; FILE REFERENCE: MTS1-P01-009  
; CURRENT APPLICATION NUMBER: US/09/862,179A  
; CURRENT FILING DATE: 2001-05-21  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 2  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PDZ inhibitory peptide  
; NAME/KEY: MISC FEATURE  
; LOCATION: (20)-(20)  
; OTHER INFORMATION: Xaa=Glu or Asp  
; NAME/KEY: MISC FEATURE  
; LOCATION: (21)-(21)  
; OTHER INFORMATION: Xaa=Met or Gly  
; NAME/KEY: MISC FEATURE  
; LOCATION: (27)-(27)  
; OTHER INFORMATION: Xaa=Ala or Pro  
US-09-862-179A-2

Query Match 5.8%; Score 141; DB 10; Length 28;  
Best Local Similarity 92.3%; Pred. No. 0.00015;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 308 CPYKVGSDYGHPIYIVQDGPQSP 333  
DB 1 CPYKVGSDYGHPIYIVQDGPQSP 26

RESULT 39  
US-10-001-887-108



```

; Sequence 108, Application US/10001887
; Patent No. US2002015564A1
; GENERAL INFORMATION:
; APPLICANT: Saiceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Racipon, Herive
; APPLICANT: Cafferey, Robert
; APPLICANT: Sun, Yongsung
; APPLICANT: Liu, Chonghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0263
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: Patent version 3.1
; SEQ ID NO 108
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-887-108

Query Match          5.7%; Score 140.5; DB 9; Length 638;
Best Local Similarity 26.4%; Pred. No. 0.0075;
Matches 97; Conservative 20; Mismatches 136; Indels 115; Gaps 18;

QY 2 GPPHSGGCGVAV-GALLLGLVGLVGLSLPEFYVNSANKRPAEGGYLYTPQIGRLDL 60
DB 208 GPP--GVDGCVGPAGALPGPQG--PSGAKGP-----GTR----- 239
QY 61 LCPHAPPPHSSPNFYLYLVGAQGRCEAPPA--PNLLTCDRPLDLRFTIKFO 118
DB 240 -----GPPGLIGPTGYC-----MPGLPGPKGRGPRAGVGLLGDGEFED----- 280
QY 119 EYSFNLWGHEFRSHHDYIATSDGTREGLSLQGGVCLTRGKVKLLRVG--QSPRGAVP 177
DB 281 -----GPPGCGGPGGLGPPGGLPGSAGLPGRRGPPPKSEAGP 318
QY 178 RKPVSFEMPERDRG---AAHSLPEGKENLPGDPTSNATSRGAEGLPPSPMVAAGAAG 233
DB 319 GGPFGVGVIGDQSPSLAKPGVPERGLPG-----AHGPPPTGPGEDGFTGRPG 371
QY 234 GALLLLGVAGAGAMCWRBRRAKPSRSRHPG--PGSFGSGSLGL--GGGGMGPR----- 286
DB 372 G-----PGVAGALG-----QKGDGLPGQPLRGPRGPGIGLQGPAGPIGPGQLPGL 417
QY 287 EAPRGELGLIRG-----GGAADPPFCPHYEKVSGDYHPYIVQDGPSPSPNITYT 339
DB 418 KGEGLPGPPEGEAGPPTGPTGPPGVGSPGITGPPGP--GPPGPGAPGAFDET 474
QY 340 SISYLEMP 347
DB 475 GIAGLHP 482

RESULT 40
US-09-823-240-10
; Sequence 10, Application US/09823240
; Patent No. US2002004813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gerler
; APPLICANT: James E. Bear
; APPLICANT: Jürgen Weiland
; APPLICANT: Joseph Loureio
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564

```

```

; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-823-240-10

Query Match          5.3%; Score 130.5; DB 10; Length 380;
Best Local Similarity 22.8%; Pred. No. 0.022;
Matches 67; Conservative 34; Mismatches 102; Indels 91; Gaps 15;

QY 35 WNSANKRPAAGGYLYVPOIGDRDLCLPRAPRPPHSSPFYERYKLYLVGAQGRCEA 94
DB 16 YDDGKRWLPAG---TGQAFSRVQT-----YHNPANSFRVV-----GRKKQ- 55
QY 95 PPAPNLLTCDRPLDLRFTIKFOEYSPN-----LWGHEFRSHHDYIATSDGTR 145
DB 56 -PDQGVVINC-----AIVRGVKNQATPNFHWMDARQVWGLNFGSKEDAAQFAA--GMA 107
QY 146 EGLSLQGGVCLTRGKVKLLRVGSPRGG---AVPRKPVSEMERDR---GAHSL- 197
DB 108 SALELEEG-----GPPPPALPTWSVPNGPSPPEVEQQKQPGPSEHIER 154
QY 198 -----PGKENLPGDPTSN--ATSRGA--EGPLPPSPMVA 229
DB 155 RVSNAGCPAPAPGAPPPPPPPPPPPPPGPPGPPGPPGPPGPPGPPGPPGPPAQAQ 214
QY 230 GAAGGALLLLGVAGAGAMCWRBRRAKPSRSRHPGSGSGSLGLGGGGM 283
DB 215 GPGGGGA---GAPGLAAAIAGVLR--KVKQEZASGGPTAPKASRGSGGGGL 263

RESULT 41
US-09-925-299-979
; Sequence 979, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: Patent ver. 2.0
; SEQ ID NO 979
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-979

Query Match          5.3%; Score 130.5; DB 10; Length 674;
Best Local Similarity 27.4%; Pred. No. 0.043;
Matches 72; Conservative 20; Mismatches 92; Indels 79; Gaps 15;

QY 141 SDGTREGLSLQGGVCLTRGKVKLLRVGQSP--RG--GAVPRKPVSEMPER--RDRGAH 194
DB 186 SRGER-GLPGVAGAV-----GPPGLIGLPGPARGPPGAVGSPGVNAPABAGRDGNPGN 240
QY 195 SLEPGKLPG-----DPTSNATSRGAEGLPP-----PSMPVAAGA 232
DB 241 DGPGRDQPGHKRGERYGNIGVGAAGAAGPGRGPGAPGAKGNRGETSPGP--VGPA 298
QY 223 GALLLLGVAGAGAMCWRBRRAKPSRSRHPG--PGSFGSGSLGL-----GGGG 281
DB 239 GA-----VGPRGSPGPGIGIKDKGPEKPGPRGLPGIKGNGLGLPGIAGHGDQAPG 353
QY 282 GWGP-----REAPGELGIA--LRG-----GGAADPPFCPHYEKVSG 316

```

Db 354 SVCPAGPRGAPSGAPAGDGTGHTGVGAPAGITGPGAGHOGAPGPPPPPPPPVSG 413  
 QY 317 ---DYGHPIYIVQDGPPOSPNNI 336  
 Db 414 GGYDFGYDGFYRADQPRSPSL 436

## RESULT 42

US-09-854-133-187  
 ; Sequence 187, Application US/09854133  
 ; Publication No. US20020183499A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Mohamath, Raedoh  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Secrist, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C10  
 ; CURRENT APPLICATION NUMBER: US/09/854,133  
 ; CURRENT FILING DATE: 2001-05-11  
 ; NUMBER OF SEQ ID NOS: 735  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 187  
 ; LENGTH: 595  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapien  
 ; US-09-854-133-187

Query Match 5.3%; Score 129; DB 9; Length 595;  
 Best Local Similarity 27.4%; Pred. No. 0.048;  
 Matches 69; Conservative 18; Mismatches 89; Indels 76; Gaps 12;

QY 85 GGAQG-RCEAPAPVLLTCDRDLRFTIKFOEYSPNLMGH----- 127  
 Db 163 GGAESPFGWEAGP-----REMGSPSGHGDGPRRRPRKRRGKG 201  
 QY 128 EFRSHDYIITATSDTRGLESLOGVCLTRGMKVLIRVQSGPRAVPRKPVSE----- 183  
 Db 202 RMGRQHEAATATATATGTAEEAGASAPES-----QAGGPRGRA--RGPRQGRRR 254  
 QY 184 MEMERDGAHSLPEKENTLPGDPTS--NATSRGAEGLPPSPMPAVAGAALALLL 240  
 Db 255 HGTORRRPPQAREEG---PRDATITLGLGTPSEGRADQSGLPLAGAAAHAAHAIIP 310  
 QY 241 G-----VAGAGMCMRRRAKPSERHPGSGFGRGSLGLGGGGMGPRAEPGEL 293  
 Db 311 GAGPAAPVGGRRGRGRRGGRGGS--AGAGGGGGRGR--GRGGGGGGG----- 357  
 QY 294 GIALRGGAADP 305  
 Db 358 GGAGRGGAAGP 369

## RESULT 43

US-09-738-973-187  
 ; Sequence 187, Application US/09738973  
 ; Patent No. US20020110563A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Lodes, Michael J.  
 ; APPLICANT: Fling, Steven P.  
 ; APPLICANT: Mohamath, Raedoh  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Secrist, Heather  
 ; APPLICANT: Indrias, Carol R.  
 ; APPLICANT: Benson, Darin R.  
 ; APPLICANT: Ellison, Mark  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Kalos, Michael D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 ; FILE REFERENCE: 210121.475C9  
 ; CURRENT APPLICATION NUMBER: US/09/738,973  
 ; CURRENT FILING DATE: 2000-12-14  
 ; NUMBER OF SEQ ID NOS: 587  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 187  
 ; LENGTH: 595  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapien  
 ; US-09-738-973-187

Query Match 5.3%; Score 129; DB 10; Length 595;  
 Best Local Similarity 27.4%; Pred. No. 0.048;  
 Matches 69; Conservative 18; Mismatches 89; Indels 76; Gaps 12;

QY 85 GGAQG-RCEAPAPVLLTCDRDLRFTIKFOEYSPNLMGH----- 127  
 Db 163 GGAESPFGWEAGP-----REMGSPSGHGDGPRRRPRKRRGKG 201  
 QY 128 EFRSHDYIITATSDTRGLESLOGVCLTRGMKVLIRVQSGPRAVPRKPVSE----- 183  
 Db 202 RMGRQHEAATATATATGTAEEAGASAPES-----QAGGPRGRA--RGPRQGRRR 254  
 QY 184 MEMERDGAHSLPEKENTLPGDPTS--NATSRGAEGLPPSPMPAVAGAALALLL 240  
 Db 255 HGTORRRPPQAREEG---PRDATITLGLGTPSEGRADQSGLPLAGAAAHAAHAIIP 310  
 QY 241 G-----VAGAGMCMRRRAKPSERHPGSGFGRGSLGLGGGGMGPRAEPGEL 293  
 Db 311 GAGPAAPVGGRRGRGRRGGRGGS--AGAGGGGGRGR--GRGGGGGG----- 357  
 QY 294 GIALRGGAADP 305  
 Db 358 GGAGRGGAAGP 369

## RESULT 44

US-09-919-497-56  
 ; Sequence 56, Application US/09919497  
 ; Patent No. US2002010662A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Multer, George L.  
 ; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
 ; FILE REFERENCE: B0801/7225  
 ; CURRENT APPLICATION NUMBER: US/09/919,497  
 ; CURRENT FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: US 60/221,735  
 ; NUMBER OF SEQ ID NOS: 100  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 56  
 ; LENGTH: 1806  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: UNSURE  
 ; LOCATION: (758)..  
 ; OTHER INFORMATION: Xaa = any amino acid  
 ; NAME/KEY: UNSURE  
 ; LOCATION: (809)..  
 ; OTHER INFORMATION: Xaa = any amino acid  
 ; US-09-919-497-56

Query Match 5.3%; Score 129; DB 10; Length 1806;  
 Best Local Similarity 29.1%; Pred. No. 0.17;  
 Matches 64; Conservative 11; Mismatches 65; Indels 80; Gaps 13;

QY 169 GSPRGAVPRKPVSEMPER-----DRGAHSLPEGKE-----NLPGD----- 206  
 Db 561 GGRGVGPPGPTGK-PGRKGRPADGGGMPGEPGAGKDRGFDGLPGLPGDKHGRG 619

QY 207 -----PTSNATSRGAGPLPPSPMPAVAGAGLALLILGVAGAGGAMCRRRAKPSB 260  
 ; SEQ ID NO 9  
 ; LENGTH: 684  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-823-240-9

Db 620 PQGPFGPDDMGKGGDEIGPGLPGEAGPRG-----LLEPGTTPGA-----PQ 665

QY 261 -----SSHPG-PGSFGRGSLGL-GGGGGMGPRAEPGELGIALRG 300  
 ; 666 PGMAVDGPPEPKNMMPGSPGPPGQGNPFGQLGPGQPIGP-----PGEKGPQGRGP 721

QY 301 -----GAADPFCHPEKVGSDYGHPIYIVODGP--PQSP 333  
 ; 722 LAGLPADGPPGHPGKEGQSGEKG-----ALGPPEPGP 755

RESULT 45  
 US-09-925-302-518  
 ; Sequence 518, Application US/09925302  
 ; Patent No. US20020044941A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA104  
 ; CURRENT APPLICATION NUMBER: US/09/925,302  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05918  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 896  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 518  
 ; LENGTH: 430  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (11)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-09-925-302-518

Query Match 5.2%; Score 127; DB 10; Length 430;  
 Best Local Similarity 29.3%; Pred. No. 0.046;  
 Matches 49; Conservative 13; Mismatches 69; Indels 36; Gaps 6;

QY 198 PGENLPGD-----PTSNATSRGAGPLPPSPMPAVAGAG--GLALLILG 241  
 ; 55 PGVFGLEKCPREGQPMGNTGPTGAVGDRGPKGDPGAPGAVGAPGIAGIPQK 114

Db 242 VAGAGCMCRRRAKPSBSRHQPGSFGSGSLGCGGGMFPRAEPGELGIALRG 301  
 ; 115 IAVDPGIVGQGRKPG--APGEMGPGGPPGEPGFRG-----APGKAGPQGRGV 163

QY 302 AADPFCHPEKVGSDYGHPIYIVODGPQPSP-----PNIYTSISV 343  
 ; 164 SAVYGF-----RGDEGPIGHQGPIGGEGAPRPSPLPGMPSVSI 206

RESULT 46  
 US-09-823-240-9  
 ; Sequence 9, Application US/09823240  
 ; Patent No. US20020044813A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Frank B. Gerler  
 ; APPLICANT: James E. Bear  
 ; APPLICANT: Jürgen Wehlend  
 ; APPLICANT: Joseph Loureiro  
 ; TITLE OF INVENTION: Methods and Products for Regulating Cell  
 ; FILE REFERENCE: M0656/7064 (HCL)  
 ; CURRENT APPLICATION NUMBER: US/09/823,240  
 ; CURRENT FILING DATE: 2001-03-30  
 ; PRIOR APPLICATION NUMBER: 60/194,564  
 ; PRIOR FILING DATE: 2000-04-03  
 ; NUMBER OF SEQ ID NOS: 11

SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 9  
 ; LENGTH: 684  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; US-09-823-240-9

Query Match 5.2%; Score 127; DB 10; Length 684;  
 Best Local Similarity 21.5%; Pred. No. 0.079;  
 Matches 82; Conservative 34; Mismatches 151; Indels 115; Gaps 15;

QY 67 PPGHSSPNEYFYKLYLVGAGRCCEAPAPNLLTCDPDLRFTIKFOEY----- 120  
 ; 234 PPAPQQQPQQQQPQQQQQQQQQVQAVGAVS-----QQYQQPHYV 275

Db 121 -----SPNLGHEPFRSHADYIYIANSDGTRESLESLOGVCITRMKYLRYGSGPRGAV 176  
 ; 276 LSNMNPNTVHOYPTQQ-----AQQCPQAPQPPPLONG-----GMVY---VGHSLPSSA 322

QY 177 PRKPYSEMPMERDGAHSLRPGKENTPG-----DPTSNATSRGAGPLPPSPMPA 227  
 ; 323 SANSTVYVASCQOMPOAHPOAPPTWPGYGGPPVPPQQAENFYGVPMFPVNP 382

QY 228 VAGAGGLALLILGVAGAGAMCRRRAKPSBSRHQPGSFGSGSLGCGG----- 281  
 ; 383 QQQQPGVPLNRMSQGGPG-----GPPAPAPPPPPSPFG-----GAAGGGPPPPAP 429

QY 282 -----GMGPREAPGELGIALRGGAADPFCPHYKVGSDYGHPIYIVODG 328  
 ; 430 PQMFGAPPPAPMGGPPAPPPAP--GGG---PPAP-----GGGAP--P 471

QY 329 PPOSPNYYTYSIVLEWPIHATITQLEFMRKSCSRVTTFLPVQVIT-----STCR 380  
 ; 472 PPPPPPLGGA-----PKKEDPQADLMGSLASQLQFLLKXKXKXVTTSPAPENSGSSTSS 524

Db 381 MTSFSEFTLNPXQACPAQWGE 402  
 ; 525 GSGGNYGTIGRSSNGMASMDE 546

RESULT 47  
 US-09-925-302-689  
 ; Sequence 689, Application US/09925302  
 ; Patent No. US20020044941A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA104  
 ; CURRENT APPLICATION NUMBER: US/09/925,302  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05918  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 896  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 689  
 ; LENGTH: 403  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (189)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-09-925-302-689

Query Match 5.2%; Score 126.5; DB 10; Length 403;  
 Best Local Similarity 30.3%; Pred. No. 0.046;  
 Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 PGENLPGDPTSNATSR-----RGAGPLPPSPMPAVAGAGLALLILGVAGAGAMC 252  
 ; 168 PGAPPPGPPAPKAGXAMGRDAGTSPGPPGVXGAG-----LQGGAPAG----- 217

QY 253 RRRAPSESRHP-----GSFGRGSLGUGGGGNGPREAREPGEIGI-----ALRGGG 301  
DB 218 ----KQAGNCTGPGGEGSKSDGSLIGPKETG---TKGKXGLGLGSGKDGKMGKDA 270  
QY 302 AADPPCPHYEKVSGDYGHFYIVODGPP 330  
DB 271 GVMGP--PQAQSGKDPGRF-----GPP 291

RESULT 48  
US-09-978-295A-614  
Sequence 614, Application US/09978295A  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gunney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kluja, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2630PIC1  
CURRENT APPLICATION NUMBER: US/09/978, 295A  
PRIOR FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796

PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083332  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083332  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084598  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085689  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697

Query Match 5.2%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.9%; Pred. No. 0.062;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;  
198 PGKENTLPGDPTSNATS-----FGAEGPLPPMPAVAGAGGATALLLTVAGAGGAMCWR 252

Db 165 PGAPPPPPPEKAKGAMGDGATGSGGPGGPPGVKGAG-----IQGPGAPG----- 214  
QY 253 RRAPPSRRHGP-----GSFRRGSLGLGGGGMGPPEAPGELI-----ALRGGG 301  
Db 215 ---KQATGTPGPGGEGSGKDDGLGPKGTEG---TKGKGDLGLPGSKGDRGMKGA 267  
QY 302 AADPFCHYEKVSDDYGHPIYIVQDGP 330  
Db 268 GVMGP--PGAQSKGDFGRP-----GPF 288  
RESULT 49  
US-09-978-697-614  
Sequence 614, Application US/09978697  
Patent No. US20020169284A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Bostein, David  
APPLICANT: Desnyere, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fors, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerlitsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Nagler, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2630PIC27  
CURRENT APPLICATION NUMBER: US/09/978,697  
PRIOR FILING DATE: 2001-10-16  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936

```

PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22

PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match      5.2%  Score 126.5; DB 9; Length 520;
Best Local Similarity 30.9%; Pred. No. 0.062;
Mismatch 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;
Qy 198 PCKENLPDPTSNATS-----RGAEGLPPSPMPAVAGAGIALLLIGVAGAGCMWR 252
```





RX MEDLINE=98370844; PubMed=9707347;  
 RA Pihlajama T., Vuorio M.M., Annunen S., Perala M., Prockop D.J.,  
 RA Ala-Kokko L.;  
 RT "Human COL9A1 and COL9A2 genes. Two genes of 90 and 15 kb code for  
 RT similar polypeptides of the same collagen molecule.";  
 RL Matrix Biol. 17:237-241(1998).  
 RL (3)  
 RP SEQUENCE FROM N.A.  
 RA Nickerson T., Harley J.;  
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 580-820 AND 835-884 FROM N.A.  
 RX MEDLINE=8937096; PubMed=2465149;  
 RA Kimura T., Mattei M.-G., Stevens J.W., Goldring M.B., Nimmiya Y.,  
 RA Olsen B.R.;  
 RT "Molecular cloning of rat and human type IX collagen cDNA and  
 RT localization of the alpha 1(IX) gene on the human chromosome 6.";  
 RL Eur. J. Biochem. 179:71-78(1989).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=90207204; PubMed=1690886;  
 RA Murasaka Y., Nishimura I., Henney A., Nimmiya Y., Olsen B.R.;  
 RT "The alpha 1 (IX) collagen gene gives rise to two different  
 RT transcripts in both mouse embryonic and human fetal RNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2400-2404(1990).  
 CC -1- FUNCTION: Structural component of hyaline cartilage and vitreous  
 CC of the eye.  
 CC -1- SUBUNIT: Heterotrimer of a alpha 1(IX), a alpha 2(IX) and a alpha  
 CC 3(IX) chain.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOPROPS; A LONG FORM (SHOWN  
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: EACH SUBUNIT IS COMPOSED OF THREE TRIPLE-HELICAL DOMAINS  
 CC INTERSPERSED WITH NONCOLLAGENOUS DOMAINS. THE GLOULAR DOMAIN AT  
 CC THE N-TERMINUS OF TYPE IX COLLAGEN MOLECULES REPRESENTS THE NC4  
 CC DOMAIN WHICH MAY PARTICIPATE IN ELECTROSTATIC INTERACTIONS WITH  
 CC POLYANIONIC GLYCOSAMINOGLYCANS IN CARTILAGE.  
 CC -1- PTM: Covalently linked to the telopeptides of type II collagen by  
 CC lysine-derived cross-links.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54412; CAA38276.1; -;  
 DR EMBL; X54413; CAA38277.1; -;  
 DR EMBL; AF036130; AAC33527.1; -;  
 DR EMBL; AF036110; AAC33527.1; JOINED.  
 DR EMBL; AF036111; AAC33527.1; JOINED.  
 DR EMBL; AF036112; AAC33527.1; JOINED.  
 DR EMBL; AF036113; AAC33527.1; JOINED.  
 DR EMBL; AF036114; AAC33527.1; JOINED.  
 DR EMBL; AF036115; AAC33527.1; JOINED.  
 DR EMBL; AF036116; AAC33527.1; JOINED.  
 DR EMBL; AF036117; AAC33527.1; JOINED.  
 DR EMBL; AF036118; AAC33527.1; JOINED.  
 DR EMBL; AF036119; AAC33527.1; JOINED.  
 DR EMBL; AF036120; AAC33527.1; JOINED.  
 DR EMBL; AF036121; AAC33527.1; JOINED.  
 DR EMBL; AF036122; AAC33527.1; JOINED.  
 DR EMBL; AF036123; AAC33527.1; JOINED.  
 DR EMBL; AF036124; AAC33527.1; JOINED.  
 DR EMBL; AF036125; AAC33527.1; JOINED.  
 DR EMBL; AF036126; AAC33527.1; JOINED.  
 DR EMBL; AF036127; AAC33527.1; JOINED.  
 DR EMBL; AF036128; AAC33527.1; JOINED.  
 DR EMBL; AF036129; AAC33527.1; JOINED.  
 DR EMBL; AF036130; AAC33528.1; JOINED.  
 DR EMBL; AF036112; AAC33528.1; JOINED.  
 DR EMBL; AF036113; AAC33528.1; JOINED.  
 DR EMBL; AF036114; AAC33528.1; JOINED.  
 DR EMBL; AF036115; AAC33528.1; JOINED.  
 DR EMBL; AF036116; AAC33528.1; JOINED.  
 DR EMBL; AF036117; AAC33528.1; JOINED.  
 DR EMBL; AF036118; AAC33528.1; JOINED.  
 DR EMBL; AF036119; AAC33528.1; JOINED.  
 DR EMBL; AF036120; AAC33528.1; JOINED.  
 DR EMBL; AF036121; AAC33528.1; JOINED.  
 DR EMBL; AF036122; AAC33528.1; JOINED.  
 DR EMBL; AF036123; AAC33528.1; JOINED.  
 DR EMBL; AF036124; AAC33528.1; JOINED.  
 DR EMBL; AF036125; AAC33528.1; JOINED.  
 DR EMBL; AF036126; AAC33528.1; JOINED.  
 DR EMBL; AF036127; AAC33528.1; JOINED.  
 DR EMBL; AF036128; AAC33528.1; JOINED.  
 DR EMBL; AF036129; AAC33528.1; JOINED.  
 DR EMBL; AF036130; AAC33528.1; JOINED.  
 DR EMBL; AF036131; AAC33528.1; JOINED.  
 DR EMBL; AF036132; AAC33528.1; JOINED.  
 DR EMBL; AF036133; AAC33528.1; JOINED.  
 DR EMBL; AF036134; AAC33528.1; JOINED.  
 DR EMBL; AF036135; AAC33528.1; JOINED.  
 DR EMBL; AF036136; AAC33528.1; JOINED.  
 DR EMBL; AF036137; AAC33528.1; JOINED.  
 DR EMBL; AF036138; AAC33528.1; JOINED.  
 DR EMBL; AF036139; AAC33528.1; JOINED.  
 DR EMBL; AF036140; AAC33528.1; JOINED.  
 DR EMBL; AF036141; AAC33528.1; JOINED.  
 DR EMBL; AF036142; AAC33528.1; JOINED.  
 DR EMBL; AF036143; AAC33528.1; JOINED.  
 DR EMBL; AF036144; AAC33528.1; JOINED.  
 DR EMBL; AF036145; AAC33528.1; JOINED.  
 DR EMBL; AF036146; AAC33528.1; JOINED.  
 DR EMBL; AF036147; AAC33528.1; JOINED.  
 DR EMBL; AF036148; AAC33528.1; JOINED.  
 DR EMBL; AF036149; AAC33528.1; JOINED.  
 DR EMBL; AF036150; AAC33528.1; JOINED.  
 DR EMBL; AF036151; AAC33528.1; JOINED.  
 DR EMBL; AF036152; AAC33528.1; JOINED.  
 DR EMBL; AF036153; AAC33528.1; JOINED.  
 DR EMBL; AF036154; AAC33528.1; JOINED.  
 DR EMBL; AF036155; AAC33528.1; JOINED.  
 DR EMBL; AF036156; AAC33528.1; JOINED.  
 DR EMBL; AF036157; AAC33528.1; JOINED.  
 DR EMBL; AF036158; AAC33528.1; JOINED.  
 DR EMBL; AF036159; AAC33528.1; JOINED.  
 DR EMBL; AF036160; AAC33528.1; JOINED.  
 DR EMBL; AF036161; AAC33528.1; JOINED.  
 DR EMBL; AF036162; AAC33528.1; JOINED.  
 DR EMBL; AF036163; AAC33528.1; JOINED.  
 DR EMBL; AF036164; AAC33528.1; JOINED.  
 DR EMBL; AF036165; AAC33528.1; JOINED.  
 DR EMBL; AF036166; AAC33528.1; JOINED.  
 DR EMBL; AF036167; AAC33528.1; JOINED.  
 DR EMBL; AF036168; AAC33528.1; JOINED.  
 DR EMBL; AF036169; AAC33528.1; JOINED.  
 DR EMBL; AF036170; AAC33528.1; JOINED.  
 DR EMBL; AF036171; AAC33528.1; JOINED.  
 DR EMBL; AF036172; AAC33528.1; JOINED.  
 DR EMBL; AF036173; AAC33528.1; JOINED.  
 DR EMBL; AF036174; AAC33528.1; JOINED.  
 DR EMBL; AF036175; AAC33528.1; JOINED.  
 DR EMBL; AF036176; AAC33528.1; JOINED.  
 DR EMBL; AF036177; AAC33528.1; JOINED.  
 DR EMBL; AF036178; AAC33528.1; JOINED.  
 DR EMBL; AF036179; AAC33528.1; JOINED.  
 DR EMBL; AF036180; AAC33528.1; JOINED.  
 DR EMBL; AF036181; AAC33528.1; JOINED.  
 DR EMBL; AF036182; AAC33528.1; JOINED.  
 DR EMBL; AF036183; AAC33528.1; JOINED.  
 DR EMBL; AF036184; AAC33528.1; JOINED.  
 DR EMBL; AF036185; AAC33528.1; JOINED.  
 DR EMBL; AF036186; AAC33528.1; JOINED.  
 DR EMBL; AF036187; AAC33528.1; JOINED.  
 DR EMBL; AF036188; AAC33528.1; JOINED.  
 DR EMBL; AF036189; AAC33528.1; JOINED.  
 DR EMBL; AF036190; AAC33528.1; JOINED.  
 DR EMBL; AF036191; AAC33528.1; JOINED.  
 DR EMBL; AF036192; AAC33528.1; JOINED.  
 DR EMBL; AF036193; AAC33528.1; JOINED.  
 DR EMBL; AF036194; AAC33528.1; JOINED.  
 DR EMBL; AF036195; AAC33528.1; JOINED.  
 DR EMBL; AF036196; AAC33528.1; JOINED.  
 DR EMBL; AF036197; AAC33528.1; JOINED.  
 DR EMBL; AF036198; AAC33528.1; JOINED.  
 DR EMBL; AF036199; AAC33528.1; JOINED.  
 DR EMBL; AF036200; AAC33528.1; JOINED.  
 DR EMBL; AF036201; AAC33528.1; JOINED.  
 DR EMBL; AF036202; AAC33528.1; JOINED.  
 DR EMBL; AF036203; AAC33528.1; JOINED.  
 DR EMBL; AF036204; AAC33528.1; JOINED.  
 DR EMBL; AF036205; AAC33528.1; JOINED.  
 DR EMBL; AF036206; AAC33528.1; JOINED.  
 DR EMBL; AF036207; AAC33528.1; JOINED.  
 DR EMBL; AF036208; AAC33528.1; JOINED.  
 DR EMBL; AF036209; AAC33528.1; JOINED.  
 DR EMBL; AF036210; AAC33528.1; JOINED.  
 DR EMBL; AF036211; AAC33528.1; JOINED.  
 DR EMBL; AF036212; AAC33528.1; JOINED.  
 DR EMBL; AF036213; AAC33528.1; JOINED.  
 DR EMBL; AF036214; AAC33528.1; JOINED.  
 DR EMBL; AF036215; AAC33528.1; JOINED.  
 DR EMBL; AF036216; AAC33528.1; JOINED.  
 DR EMBL; AF036217; AAC33528.1; JOINED.  
 DR EMBL; AF036218; AAC33528.1; JOINED.  
 DR EMBL; AF036219; AAC33528.1; JOINED.  
 DR EMBL; AF036220; AAC33528.1; JOINED.  
 DR EMBL; AF036221; AAC33528.1; JOINED.  
 DR EMBL; AF036222; AAC33528.1; JOINED.  
 DR EMBL; AF036223; AAC33528.1; JOINED.  
 DR EMBL; AF036224; AAC33528.1; JOINED.  
 DR EMBL; AF036225; AAC33528.1; JOINED.  
 DR EMBL; AF036226; AAC33528.1; JOINED.  
 DR EMBL; AF036227; AAC33528.1; JOINED.  
 DR EMBL; AF036228; AAC33528.1; JOINED.  
 DR EMBL; AF036229; AAC33528.1; JOINED.  
 DR EMBL; AF036230; AAC33528.1; JOINED.  
 DR EMBL; AF036231; AAC33528.1; JOINED.  
 DR EMBL; AF036232; AAC33528.1; JOINED.  
 DR EMBL; AF036233; AAC33528.1; JOINED.  
 DR EMBL; AF036234; AAC33528.1; JOINED.  
 DR EMBL; AF036235; AAC33528.1; JOINED.  
 DR EMBL; AF036236; AAC33528.1; JOINED.  
 DR EMBL; AF036237; AAC33528.1; JOINED.  
 DR EMBL; AF036238; AAC33528.1; JOINED.  
 DR EMBL; AF036239; AAC33528.1; JOINED.  
 DR EMBL; AF036240; AAC33528.1; JOINED.  
 DR EMBL; AF036241; AAC33528.1; JOINED.  
 DR EMBL; AF036242; AAC33528.1; JOINED.  
 DR EMBL; AF036243; AAC33528.1; JOINED.  
 DR EMBL; AF036244; AAC33528.1; JOINED.  
 DR EMBL; AF036245; AAC33528.1; JOINED.  
 DR EMBL; AF036246; AAC33528.1; JOINED.  
 DR EMBL; AF036247; AAC33528.1; JOINED.  
 DR EMBL; AF036248; AAC33528.1; JOINED.  
 DR EMBL; AF036249; AAC33528.1; JOINED.  
 DR EMBL; AF036250; AAC33528.1; JOINED.  
 DR EMBL; AF036251; AAC33528.1; JOINED.  
 DR EMBL; AF036252; AAC33528.1; JOINED.  
 DR EMBL; AF036253; AAC33528.1; JOINED.  
 DR EMBL; AF036254; AAC33528.1; JOINED.  
 DR EMBL; AF036255; AAC33528.1; JOINED.  
 DR EMBL; AF036256; AAC33528.1; JOINED.  
 DR EMBL; AF036257; AAC33528.1; JOINED.  
 DR EMBL; AF036258; AAC33528.1; JOINED.  
 DR EMBL; AF036259; AAC33528.1; JOINED.  
 DR EMBL; AF036260; AAC33528.1; JOINED.  
 DR EMBL; AF036261; AAC33528.1; JOINED.  
 DR EMBL; AF036262; AAC33528.1; JOINED.  
 DR EMBL; AF036263; AAC33528.1; JOINED.  
 DR EMBL; AF036264; AAC33528.1; JOINED.  
 DR EMBL; AF036265; AAC33528.1; JOINED.  
 DR EMBL; AF036266; AAC33528.1; JOINED.  
 DR EMBL; AF036267; AAC33528.1; JOINED.  
 DR EMBL; AF036268; AAC33528.1; JOINED.  
 DR EMBL; AF036269; AAC33528.1; JOINED.  
 DR EMBL; AF036270; AAC33528.1; JOINED.  
 DR EMBL; AF036271; AAC33528.1; JOINED.  
 DR EMBL; AF036272; AAC33528.1; JOINED.  
 DR EMBL; AF036273; AAC33528.1; JOINED.  
 DR EMBL; AF036274; AAC33528.1; JOINED.  
 DR EMBL; AF036275; AAC33528.1; JOINED.  
 DR EMBL; AF036276; AAC33528.1; JOINED.  
 DR EMBL; AF036277; AAC33528.1; JOINED.  
 DR EMBL; AF036278; AAC33528.1; JOINED.  
 DR EMBL; AF036279; AAC33528.1; JOINED.  
 DR EMBL; AF036280; AAC33528.1; JOINED.  
 DR EMBL; AF036281; AAC33528.1; JOINED.  
 DR EMBL; AF036282; AAC33528.1; JOINED.  
 DR EMBL; AF036283; AAC33528.1; JOINED.  
 DR EMBL; AF036284; AAC33528.1; JOINED.  
 DR EMBL; AF036285; AAC33528.1; JOINED.  
 DR EMBL; AF036286; AAC33528.1; JOINED.  
 DR EMBL; AF036287; AAC33528.1; JOINED.  
 DR EMBL; AF036288; AAC33528.1; JOINED.  
 DR EMBL; AF036289; AAC33528.1; JOINED.  
 DR EMBL; AF036290; AAC33528.1; JOINED.  
 DR EMBL; AF036291; AAC33528.1; JOINED.  
 DR EMBL; AF036292; AAC33528.1; JOINED.  
 DR EMBL; AF036293; AAC33528.1; JOINED.  
 DR EMBL; AF036294; AAC33528.1; JOINED.  
 DR EMBL; AF036295; AAC33528.1; JOINED.  
 DR EMBL; AF036296; AAC33528.1; JOINED.  
 DR EMBL; AF036297; AAC33528.1; JOINED.  
 DR EMBL; AF036298; AAC33528.1; JOINED.  
 DR EMBL; AF036299; AAC33528.1; JOINED.  
 DR EMBL; AF036300; AAC33528.1; JOINED.  
 DR EMBL; AF036301; AAC33528.1; JOINED.  
 DR EMBL; AF036302; AAC33528.1; JOINED.  
 DR EMBL; AF036303; AAC33528.1; JOINED.  
 DR EMBL; AF036304; AAC33528.1; JOINED.  
 DR EMBL; AF036305; AAC33528.1; JOINED.  
 DR EMBL; AF036306; AAC33528.1; JOINED.  
 DR EMBL; AF036307; AAC33528.1; JOINED.  
 DR EMBL; AF036308; AAC33528.1; JOINED.  
 DR EMBL; AF036309; AAC33528.1; JOINED.  
 DR EMBL; AF036310; AAC33528.1; JOINED.  
 DR EMBL; AF036311; AAC33528.1; JOINED.  
 DR EMBL; AF036312; AAC33528.1; JOINED.  
 DR EMBL; AF036313; AAC33528.1; JOINED.  
 DR EMBL; AF036314; AAC33528.1; JOINED.  
 DR EMBL; AF036315; AAC33528.1; JOINED.  
 DR EMBL; AF036316; AAC33528.1; JOINED.  
 DR EMBL; AF036317; AAC33528.1; JOINED.  
 DR EMBL; AF036318; AAC33528.1; JOINED.  
 DR EMBL; AF036319; AAC33528.1; JOINED.  
 DR EMBL; AF036320; AAC33528.1; JOINED.  
 DR EMBL; AF036321; AAC33528.1; JOINED.  
 DR EMBL; AF036322; AAC33528.1; JOINED.  
 DR EMBL; AF036323; AAC33528.1; JOINED.  
 DR EMBL; AF036324; AAC33528.1; JOINED.  
 DR EMBL; AF036325; AAC33528.1; JOINED.  
 DR EMBL; AF036326; AAC33528.1; JOINED.  
 DR EMBL; AF036327; AAC33528.1; JOINED.  
 DR EMBL; AF036328; AAC33528.1; JOINED.  
 DR EMBL; AF036329; AAC33528.1; JOINED.  
 DR EMBL; AF036330; AAC33528.1; JOINED.  
 DR EMBL; AF036331; AAC33528.1; JOINED.  
 DR EMBL; AF036332; AAC33528.1; JOINED.  
 DR EMBL; AF036333; AAC33528.1; JOINED.  
 DR EMBL; AF036334; AAC33528.1; JOINED.  
 DR EMBL; AF036335; AAC33528.1; JOINED.  
 DR EMBL; AF036336; AAC33528.1; JOINED.  
 DR EMBL; AF036337; AAC33528.1; JOINED.  
 DR EMBL; AF036338; AAC33528.1; JOINED.  
 DR EMBL; AF036339; AAC33528.1; JOINED.  
 DR EMBL; AF036340; AAC33528.1; JOINED.  
 DR EMBL; AF036341; AAC33528.1; JOINED.  
 DR EMBL; AF036342; AAC33528.1; JOINED.  
 DR EMBL; AF036343; AAC33528.1; JOINED.  
 DR EMBL; AF036344; AAC33528.1; JOINED.  
 DR EMBL; AF036345; AAC33528.1; JOINED.  
 DR EMBL; AF036346; AAC33528.1; JOINED.  
 DR EMBL; AF036347; AAC33528.1; JOINED.  
 DR EMBL; AF036348; AAC33528.1; JOINED.  
 DR EMBL; AF036349; AAC33528.1; JOINED.  
 DR EMBL; AF036350; AAC33528.1; JOINED.  
 DR EMBL; AF036351; AAC33528.1; JOINED.  
 DR EMBL; AF036352; AAC33528.1; JOINED.  
 DR EMBL; AF036353; AAC33528.1; JOINED.  
 DR EMBL; AF036354; AAC33528.1; JOINED.  
 DR EMBL; AF036355; AAC33528.1; JOINED.  
 DR EMBL; AF036356; AAC33528.1; JOINED.  
 DR EMBL; AF036357; AAC33528.1; JOINED.  
 DR EMBL; AF036358; AAC33528.1; JOINED.  
 DR EMBL; AF036359; AAC33528.1; JOINED.  
 DR EMBL; AF036360; AAC33528.1; JOINED.  
 DR EMBL; AF036361; AAC33528.1; JOINED.  
 DR EMBL; AF036362; AAC33528.1; JOINED.  
 DR EMBL; AF036363; AAC33528.1; JOINED.  
 DR EMBL; AF036364; AAC33528.1; JOINED.  
 DR EMBL; AF036365; AAC33528.1; JOINED.  
 DR EMBL; AF036366; AAC33528.1; JOINED.  
 DR EMBL; AF036367; AAC33528.1; JOINED.  
 DR EMBL; AF036368; AAC33528.1; JOINED.  
 DR EMBL; AF036369; AAC33528.1; JOINED.  
 DR EMBL; AF036370; AAC33528.1; JOINED.  
 DR EMBL; AF036371; AAC33528.1; JOINED.  
 DR EMBL; AF036372; AAC33528.1; JOINED.  
 DR EMBL; AF036373; AAC33528.1; JOINED.  
 DR EMBL; AF036374; AAC33528.1; JOINED.  
 DR EMBL; AF036375; AAC33528.1; JOINED.  
 DR EMBL; AF036376; AAC33528.1; JOINED.  
 DR EMBL; AF036377; AAC33528.1; JOINED.  
 DR EMBL; AF036378; AAC33528.1; JOINED.  
 DR EMBL; AF036379; AAC33528.1; JOINED.  
 DR EMBL; AF036380; AAC33528.1; JOINED.  
 DR EMBL; AF036381; AAC33528.1; JOINED.  
 DR EMBL; AF036382; AAC33528.1; JOINED.  
 DR EMBL; AF036383; AAC33528.1; JOINED.  
 DR EMBL; AF036384; AAC33528.1; JOINED.  
 DR EMBL; AF036385; AAC33528.1; JOINED.  
 DR EMBL; AF036386; AAC33528.1; JOINED.  
 DR EMBL; AF036387; AAC33528.1; JOINED.  
 DR EMBL; AF036388; AAC33528.1; JOINED.  
 DR EMBL; AF036389; AAC33528.1; JOINED.  
 DR EMBL; AF036390; AAC33528.1; JOINED.  
 DR EMBL; AF036391; AAC33528.1; JOINED.  
 DR EMBL; AF036392; AAC33528.1; JOINED.  
 DR EMBL; AF036393; AAC33528.1; JOINED.  
 DR EMBL; AF036394; AAC33528.1; JOINED.  
 DR EMBL; AF036395; AAC33528.1; JOINED.  
 DR EMBL; AF036396; AAC33528.1; JOINED.  
 DR EMBL; AF036397; AAC33528.1; JOINED.  
 DR EMBL; AF036398; AAC33528.1; JOINED.  
 DR EMBL; AF036399; AAC33528.1; JOINED.  
 DR EMBL; AF036400; AAC33528.1; JOINED.  
 DR EMBL; AF036401; AAC33528.1; JOINED.  
 DR EMBL; AF036402; AAC33528.1; JOINED.  
 DR EMBL; AF036403; AAC33528.1; JOINED.  
 DR EMBL; AF036404; AAC33528.1; JOINED.  
 DR EMBL; AF036405; AAC33528.1; JOINED.  
 DR EMBL; AF036406; AAC33528.1; JOINED.  
 DR EMBL; AF036407; AAC33528.1; JOINED.  
 DR EMBL; AF036408; AAC33528.1; JOINED.  
 DR EMBL; AF036409; AAC33528.1; JOINED.  
 DR EMBL; AF036410; AAC33528.1; JOINED.  
 DR EMBL; AF036411; AAC33528.1; JOINED.  
 DR EMBL; AF036412; AAC33528.1; JOINED.  
 DR EMBL; AF036413; AAC33528.1; JOINED.  
 DR EMBL; AF036414; AAC33528.1; JOINED.  
 DR EMBL; AF036415; AAC33528.1; JOINED.  
 DR EMBL; AF036416; AAC33528.1; JOINED.  
 DR EMBL; AF036417; AAC33528.1; JOINED.  
 DR EMBL; AF036418; AAC33528.1; JOINED.  
 DR EMBL; AF036419; AAC33528.1; JOINED.  
 DR EMBL; AF036420; AAC33528.1; JOINED.  
 DR EMBL; AF036421; AAC33528.1; JOINED.  
 DR EMBL; AF036422; AAC33528.1; JOINED.  
 DR EMBL; AF036423; AAC33528.1; JOINED.  
 DR EMBL; AF036424; AAC33528.1; JOINED.  
 DR EMBL; AF036425; AAC33528.1; JOINED.  
 DR EMBL; AF036426; AAC33528.1; JOINED.  
 DR EMBL; AF036427; AAC33528.1; JOINED.  
 DR EMBL; AF036428; AAC33528.1; JOINED.  
 DR EMBL; AF036429; AAC33528.1; JOINED.  
 DR EMBL; AF036430; AAC33528.1; JOINED.  
 DR EMBL; AF036431; AAC33528.1; JOINED.  
 DR EMBL; AF036432; AAC33528.1; JOINED.  
 DR EMBL; AF036433; AAC33528.1; JOINED.  
 DR EMBL; AF036434; AAC33528.1; JOINED.  
 DR EMBL; AF036435; AAC33528.1; JOINED.  
 DR EMBL; AF036436; AAC33528.1; JOINED.  
 DR EMBL; AF036437; AAC33528.1; JOINED.  
 DR EMBL; AF036438; AAC33528.1; JOINED.  
 DR EMBL; AF036439; AAC33528.1; JOINED.  
 DR EMBL; AF036440; AAC33528.1; JOINED.  
 DR EMBL; AF036441; AAC33528.1; JOINED.  
 DR EMBL; AF036442; AAC33528.1; JOINED.  
 DR EMBL; AF036443; AAC33528.1; JOINED.  
 DR EMBL; AF036444; AAC33528.1; JOINED.  
 DR EMBL; AF036445; AAC33528.1; JOINED.  
 DR EMBL; AF036446; AAC33528.1; JOINED.  
 DR EMBL; AF036447; AAC33528.1; JOINED.  
 DR EMBL; AF036448; AAC33528.1; JOINED.  
 DR EMBL; AF036449; AAC33528.1; JOINED.  
 DR EMBL; AF036450; AAC33528.1; JOINED.  
 DR EMBL; AF036451; AAC33528.1; JOINED.  
 DR EMBL; AF036452; AAC33528.1; JOINED.  
 DR EMBL; AF036453; AAC33528.1; JOINED.  
 DR EMBL; AF036454; AAC33528.1; JOINED.  
 DR EMBL; AF036455; AAC33528.1; JOINED.  
 DR EMBL; AF036456; AAC33528.1; JOINED.  
 DR EMBL; AF036457; AAC33528.1; JOINED.  
 DR EMBL; AF036458; AAC33528.1; JOINED.  
 DR EMBL; AF036459; AAC33528.1; JOINED.  
 DR EMBL; AF036460; AAC33528.1; JOINED.  
 DR EMBL; AF036461; AAC33528.1; JOINED.  
 DR EMBL; AF036462; AAC33528.1; JOINED.  
 DR EMBL; AF036463; AAC33528.1; JOINED.  
 DR EMBL; AF036464; AAC33528.1; JOINED.  
 DR EMBL; AF036465; AAC33528.1; JOINED.  
 DR EMBL; AF036466; AAC33528.1; JOINED.  
 DR EMBL; AF036467; AAC33528.1; JOINED.  
 DR EMBL; AF036468; AAC33528.1; JOINED.  
 DR EMBL; AF036469; AAC33528.1; JOINED.  
 DR EMBL; AF036470; AAC33528.1; JOINED.  
 DR EMBL; AF036471; AAC33528.



RP SEQUENCE FROM N.A.  
 RX MEDLINE=91210992; PubMed=2019595;  
 RA Murgaki Y., Jaccenko O., Apte S., Mattei M.-G., Nishimura Y.,  
 RA Olsen B.R.;  
 RT "The alpha 2(VIII) collagen gene. A novel member of the short chain  
 RT collagen family located on the human chromosome 1.";  
 RL J. Biol. Chem. 266:7721-7727(1991).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT  
 CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.  
 CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION  
 CC WITH ALPHA 1(VIII) TYPE COLLAGENS.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 1 TYPES VIII AND X COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 CC EMBL; M60832; AAA62822.1; -  
 CC Gene; HGNC:2216; COL8A2.  
 DR MIM; 120252;  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF01391; Collagen; 8.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS00113; C1Q; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen.  
 FT NON\_TER 1  
 FT DOMAIN <1 11 NONHELICAL REGION (NC2).  
 FT DOMAIN 12 468 TRIPLE-HELICAL REGION.  
 FT DOMAIN 469 635 NONHELICAL REGION (NC1).  
 FT DOMAIN 500 635 C1Q.  
 SQ SEQUENCE 635 AA; 60527 MW; 22A261164754F771 CRC64;  
 Query Match 5.8%; Score 142.5; DB 1; Length 635;  
 Best Local Similarity 26.1%; Pred. No. 0.02;  
 Matches 74; Conservative 17; Mismatches 85; Indels 107; Gaps 14;  
 QY 147 GLESLG-----GVCILTRMKVILRVGQSPRGAVP-----RKPVSEMEMEDRGAA 193  
 DB 222 GLPGPGPGGAKGEPGTRPGLI-----GPTYGMFGLPGPRGDRGPGVPLGDRG-- 275  
 QY 194 HSLPEKENTLPGD-----PTSNATS-----RGAEGPLPPSPMPAVAGAAG 233  
 DB 276 ---EPGDDDDPREGQPGQIGCPGPGAGLPGRRGPPGLRGAAPGCPGVPGRGNGG 332  
 QY 234 ---GLALLLIGVAGAGAMCRRRRAPSESRRHP-----GSFGRG 272  
 DB 333 PSLIA-----GKGVGP-----ERGLPGARCPGPTGKPGEGFTGRDGPVAGALGQX 382  
 QY 273 GSIGL-----GGGGGMGR-----EAEGELGILARG-----GGAAD 304  
 DB 383 GDLGLPGGGLGPGSGIPGLQGPAGPIPGQIPGLKGEPLGPPGEGRAGEPGTAGRG 442  
 QY 305 PPECPIYKVSQDGHVYIVDGGPQSPQPNITYTSISVLEMP 347  
 DB 443 PGVPGSPGITGPPGLP-----GPGARGAGFDEGTIGLHLR 479

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 2(V) chain precursor.  
 GN COL5A2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-463 FROM N.A.  
 RX MEDLINE=89123368; PubMed=2914927;  
 RA Woodbury D., Benson-Chanda V., Ramirez F.;  
 RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms  
 RT to the structural criteria of a fibrillar procollagen molecule.";  
 RL J. Biol. Chem. 264:2735-2738(1989).  
 RN [2]  
 RP SEQUENCE OF 398-1496 FROM N.A.  
 RX MEDLINE=87146331; PubMed=3029669;  
 RA Weil D., Bernard M.P., Gargano S., Ramirez F.;  
 RT "The pro alpha 2(V) collagen gene is evolutionarily related to the  
 RT major fibrillar-forming collagens.";  
 RL Nucleic Acids Res. 15:181-198(1987).  
 RN [3]  
 RP SEQUENCE OF 1227-1496 FROM N.A.  
 RX MEDLINE=85289337; PubMed=2411731;  
 RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;  
 RT "Complete primary structure of the human alpha 2 type V procollagen  
 RT COOH-terminal propeptide.";  
 RL J. Biol. Chem. 260:11216-11222(1985).  
 RN [4]  
 RP SEQUENCE OF 1449-1496 FROM N.A.  
 RX MEDLINE=89138450; PubMed=3224983;  
 RA Tzipouras P., Schwartz R.C., Liddell A.C., Salkeid C.S., Weil D.;  
 RA Ramirez F.;  
 RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,  
 RT located on the long arm of human chromosome 2.";  
 RL Genomics 3:275-277(1988).  
 RN [5]  
 RP SEQUENCE OF 208-227.  
 RC TISSUE=Placenta;  
 RX MEDLINE=82239022; PubMed=1571108;  
 RA Mann K.;  
 RT "Isolation of the alpha 3-chain of human type V collagen and  
 RT characterization by partial sequencing.";  
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).  
 RN [6]  
 RP SEQUENCE OF 288-297 AND 606-617.  
 RC TISSUE=Bone;  
 RX MEDLINE=94237164; PubMed=8181482;  
 RA Moradi-Ameli M., Rousseau J.C., Klemen J.P., Champliand M.F.,  
 RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;  
 RT "Diversity in the processing events at the N-terminus of type-V  
 RT collagen.";  
 RL Eur. J. Biochem. 221:987-995(1994).  
 RN [7]  
 RP DISEASE.  
 RX PubMed=9425231;  
 RA Michalikova K., Susic M., Willing M.C., Wenzstrup R.J., Cole W.G.;  
 RT "Mutations of the alpha2(V) chain of type V collagen impair matrix  
 RT assembly and produce Ehlers-Danlos syndrome type I.";  
 RL Hum. Mol. Genet. 7:249-255(1998).  
 RN [8]  
 RP VARIANT EDS-II ARG-960.  
 RX PubMed=9783710;  
 RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,  
 RA Burrows N.P.;  
 RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type  
 RT II.";  
 RL J. Med. Genet. 35:846-848(1998).  
 CC -1- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE  
 CC COMPONENT OF NEARLY UBICUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS  
 CC TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.  
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN

CC MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND  
 CC ONE ALPHA 3(V) CHAINS IN PLACENTA.  
 CC -1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome  
 CC gravis or type I (EDS-I), and Ehlers-Danlos syndrome mitis or type  
 CC II (EDS-II). Ehlers-Danlos syndrome is a genetically and  
 CC phenotypically heterogeneous connective-tissue disorder  
 CC characterized by loose-jointedness and fragile, velvety,  
 CC stretchable, bruisable skin that heals with peculiar 'cigarette-  
 CC paper' scars. EDS-I and EDS-II are autosomal dominant traits.  
 CC -1- SIMILARITY: CONTAINS 1 WFCC DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J04478; AAA51959.1; -  
 DR EMBL: X04758; CAA28454.1; -  
 DR EMBL: M11718; AAA52058.1; -  
 DR PIR: A25374; A25374.  
 DR PIR: A25874; A25874.  
 DR PIR: A30017; A30017.  
 DR PIR: A31427; A31427.  
 DR Genew; HGNC:2210; COL5A2.  
 DR MIM: 120190; -  
 DR MIM: 130000; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib collagen\_C.  
 DR InterPro: IPR001007; WFCC.  
 DR Pfam: PF00093; VWC; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Prodom; PD00007; Collagen; 5.  
 DR Prodom; PD002078; Fib collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; WFCC; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;  
 KM Disease mutation.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1226 COLLAGEN ALPHA 2(V) CHAIN.  
 FT PROPEP 1227 1496 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 39 97 WFCC.  
 FT MOD\_RES 230 290 HYDROXYLATION.  
 FT MOD\_RES 293 293 HYDROXYLATION.  
 FT MOD\_RES 296 296 HYDROXYLATION.  
 FT MOD\_RES 608 608 HYDROXYLATION.  
 FT MOD\_RES 614 614 HYDROXYLATION.  
 FT VARIANT 960 960 G -> R (IN EDS-II).  
 FT VARIANT 960 960 /FTID=VAR\_013568.  
 FT CONFLICT 292 292 A -> P (IN REF. 6).  
 FT CONFLICT 1418 1418 K -> T (IN REF. 3).  
 FT CONFLICT 1438 1438 F -> S (IN REF. 3).  
 FT CONFLICT 1460 1460 E -> Q (IN REF. 4).  
 FT CONFLICT 1496 1496 V -> A (IN REF. 4).  
 FT CONFLICT 1496 1496 V -> A (IN REF. 4).  
 SQ SEQUENCE 1496 AA, 144720 MW, 82827C17A8644F5A CRC64;  
 Query Match 5.8%; Score 142.5; DB 1; Length 1496;  
 Best Local Similarity 29.1%; Pred. No. 0.048;  
 Matches 55; Conservative 15; Mismatches 76; Indels 43; Gaps 9;

DB 555 GDRGGRGRLPARG-----LTGNVGQPGKXGLGAPEDDGRPGPSGIKIGQGC 609  
 QY 274 SLGLGGGGNGPPEAPPELGLALRGGAADPPFCHEKVGDIYIYQDGP--- 330  
 DB 610 TMDLPGRKSGNGDPGRGENG-----NPGVQGRGAPKDGKVGYPGP-----GPGILR 659  
 QY 331 -----QSP 334  
 DB 660 GERGGGPP 668  
 RESULT 40  
 CA18 HUMAN STANDARD; PRT; 744 AA.  
 ID CA18 HUMAN STANDARD; PRT; 744 AA.  
 AC P27658; Q96D07; -  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).  
 GN COL8A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91231001; PubMed=2029894;  
 RA Muragaki Y., Mattei M.-G., Yamaguchi N., Olsen B.R., Ninomiya Y.;  
 RT "The complete primary structure of the human alpha 1 (VIII) chain and  
 RT assignment of its gene (COL8A1) to chromosome 3.";  
 RL Eur. J. Biochem. 197;615-622 (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Straubeberg R.;  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DDBJ databases.  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT  
 CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.  
 CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION  
 CC WITH ALPHA 2(VIII) TYPE COLLAGENS.  
 CC -1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPEPTIDES ARE PRESENT  
 CC AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE  
 CC THE HIGH THERMAL STABILITY OF THIS REGION.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X57527; CAA0748.1; -  
 DR EMBL: BC013581; AAH13581.1; -  
 DR PIR: S15435; S15435.  
 DR Genew; HGNC:2215; COL8A1.  
 DR MIM: 120251; -  
 DR InterPro: IPR001073; C1G.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF00386; C1G; 1.  
 DR Pfam: PF01391; Collagen; 8.  
 DR Prodom; PD00007; COMPLEMENTC1G.  
 DR Prodom; PD000007; Collagen; 1.  
 DR SMART; SMO0110; C1G; 1.  
 DR PROSITE; PS01113; C1G; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Cell adhesion; Collagen; Signal.  
 FT SIGNAL 1 28

```

FT CHAIN 29 744 COLLAGEN ALPHA 1(VI) CHAIN.
FT DOMAIN 29 117 NONHELICAL REGION (NC2).
FT DOMAIN 118 571 TRIPLE-HELICAL REGION (COLL).
FT DOMAIN 572 744 NONHELICAL REGION (NC1).
FT DOMAIN 609 744 CLO.
FT CONFLICT 262 744 P -> L (IN REF. 1).
FT CONFLICT 297 744 P -> R (IN REF. 1).
FT CONFLICT 344 744 P -> A (IN REF. 1).
FT CONFLICT 382 744 A -> S (IN REF. 1).
FT CONFLICT 388 744 P -> S (IN REF. 1).
FT CONFLICT 454 744 L -> F (IN REF. 1).
FT CONFLICT 464 744 A -> H (IN REF. 1).
FT CONFLICT 601 744 Y -> T (IN REF. 1).
FT CONFLICT 631 744 A -> G (IN REF. 1).
SQ SEQUENCE 744 AA; 73364 MW; 28C1B0955DE2C9A3 CRC64;

Query Match 5.8%; Score 142; DB 1; Length 744;
Best Local Similarity 23.3%; Pred. No. 0.025; Indels 122; Gaps 13;
Matches 72; Conservative 16; Mismatches 99;

QY 67 PPGHSSPNYEFYKLVGGAQGRCEAPPAENLLTCDPDLDFRTIKFQEVSPNIMG 126
DB 198 PPGHGLPG-----IGKPGGPGLPQPGPK----- 222
QY 127 HEFRSHHDYITATSDGTREGLESIGQVCLTRGMKVLNFGQSPRGAAPRKVSEKPM 186
DB 223 -----GDRPKGKLPQPG-----LRGKDGKKGCMGAPGVKGP- 256
QY 187 ERDRGAHSLERKENTLPDPTSNMYS--RGAEGLPPSPMPVAVAGAAGLLLL-----G 241
DB 257 -----PDMHG-PGPGVGLPVGKPGVGTGPPGPGPLGKGAPEBPQPGPIGVQSGP 311
QY 242 VVAGAGAMCMRRRRAPSPSESRHPG-----PGSGRGGS 274
DB 312 IPIGIG-----KPGGDGI.PGPGFPGKGEQGLPGLPGLPGIKRPFPGPKGD 361
QY 275 LGLGG-GGGMGPRERAPGLGIALGGGADBPFCHEKXSGDGHVYIVQDP----- 329
DB 362 RKGGVPGALGPR-GEKPGTGA.PGIGG-----PGEGLPGLTGPMPGPAIGFPQKGE 416
QY 330 ----PQSGP 334
DB 417 GIVPGGPP 425

RESULT 41
CA26_MOUSE STANDARD; PRT; 1029 AA.
ID CA26_MOUSE
AC 002788; 005505;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(VI) chain precursor.
GN COL6A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RN [1] SEQUENCE FROM N.A.
RA Ibrahim A., Barton S., Dani C.;
RA Submitted (MAY-1992) to the EMBL/Genbank/DBJ databases.
RX MEDLINE=93143859; PubMed=8380980;
RA Ibrahim A., Bertrand B., Barton S., Amri E.Z., Grimaldi P.,
RA "Cloning of alpha 2 chain of type VI collagen and expression during
RT mouse development.";
RT Biochem. J. 289:141-147(1993).
RN [3]
RN [3] SEQUENCE OF 266-600 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Fibroblast;

```

```

RX MEDLINE=91226374; PubMed=1709252;
RA Constantinou C.D., Jimenez S.A.;
RT "Structure of cDNAs encoding the triple-helical domain of murine
RT alpha 2 (VI) collagen chain and comparison to human and chick
RT homologues. Use of polymerase chain reaction and partially degenerate
RT Oligonucleotide for generation of novel cDNA clones.";
RT Matrix 11:1-9(1991).
RN [4]
RP SEQUENCE OF 659-1029 FROM N.A.
RX MEDLINE=93256888; PubMed=8489506;
RA Zhang R.Z., Pan T.C., Timpl R., Chu M.-L.;
RT "Cloning and sequence analysis of cDNAs encoding the alpha 1, alpha 2
RT and alpha 3 chains of mouse collagen VI.";
RT Biochem. J. 291:787-792(1993).
CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),
CC ALPHA 2(VI), AND ALPHA 3(VI).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADIPOSE TISSUE, LUNG,
CC ADRENAL GLANDS AND OVARY. LOWER LEVELS IN TESTIS, TONGUE, SKIN,
CC KIDNEY, HEART, INTESTINE AND SPLEEN. NO EXPRESSION IN SKELETAL
CC MUSCLE OR LIVER.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 3 VMPA DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X65582; CAA46541.1; -
DR EMBL; X62332; CAA44206.1; -
DR EMBL; Z18272; CAA79153.1; -
DR EMBL; I06343; AAA79441.1; -
DR F01; S13745; S13745.
DR W01; M01; 88460; Col6a2.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 3.
DR Pfam; PF01391; Collagen; 4.
DR PRINTS; PRO0453; VMPADOMAIN.
DR SMART; SM00327; VWA; 3.
DR PROSITE; PS50234; VMPA; 3.
KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Glycoprotein; Cell adhesion; Collagen; Signal.
FT CHAIN 1 20
FT DOMAIN 21 1029
FT DOMAIN 21 265
FT DOMAIN 266 600
FT DOMAIN 601 1029
FT DOMAIN 54 178
FT DOMAIN 623 748
FT DOMAIN 843 965
FT SITE 376 378
FT SITE 436 438
FT SITE 499 501
FT SITE 508 510
FT SITE 549 551
FT CARBOHYD 150 150
FT CARBOHYD 337 337
FT CARBOHYD 640 640
FT CARBOHYD 795 795
FT CARBOHYD 907 907
FT CARBOHYD 963 963
FT CONFLICT 268 268
FT CONFLICT 804 804
FT CONFLICT 848 848
FT CONFLICT 962 966
FT CONFLICT 976 977
SQ SEQUENCE 1029 AA; 109811 MW; EDALBA0AF266F5D CRC64;

```



DE Collagen alpha 1(III) chain precursor.  
 GN COL3A1.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RC TISSUE=Skin fibroblast;  
 RX MEDLINE=89350838; PubMed=2764886;  
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuitaniemi H.,  
 RA Prockop D.J.;  
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
 RT chain of human type III procollagen. Differences in protein structure  
 RT from type I procollagen and conservation of codon preferences.";  
 RL Biochem. J. 260:1509-1516(1989).  
 RN [2]  
 RP SEQUENCE OF 149-1225 FROM N.A.  
 RX MEDLINE=89386015; PubMed=2780304;  
 RA Janeczko R.A., Ramirez F.;  
 RT "Nucleotide and amino acid sequences of the entire human alpha 1  
 RT (III) collagen.";  
 RL Nucleic Acids Res. 17:6742-6742(1989).  
 RN [3]  
 RP SEQUENCE OF 168-398.  
 RX MEDLINE=77134724; PubMed=557335;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of cyanogen  
 RT bromide peptides from the amino-terminal segment of type III collagen  
 RT of human liver.";  
 RL Biochemistry 16:1158-1164(1977).  
 RN [4]  
 RP REVISIONS.  
 RA Seyer J.M.;  
 RT Submitted (DEC-1977) to the PIR data bank.  
 RL [5]  
 RP SEQUENCE OF 399-727.  
 RX MEDLINE=79000343; PubMed=687591;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of five  
 RT consecutive CNBr peptides from type III collagen of human liver.";  
 RL Biochemistry 17:13404-13411(1978).  
 RN [6]  
 RP SEQUENCE OF 728-964.  
 RX MEDLINE=80198282; PubMed=6246925;  
 RA Seyer J.M., Mainardi C., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha 1  
 RT (III)-CBs from type III collagen of human liver.";  
 RL Biochemistry 19:1583-1589(1980).  
 RN [7]  
 RP SEQUENCE OF 950-1466 FROM N.A.  
 RX MEDLINE=88189837; PubMed=3357782;  
 RA Mankoo B.S., Dalglish R.;  
 RT "Human pro. alpha 1(III) collagen: cDNA sequence for the 3' end.";  
 RL Nucleic Acids Res. 16:2337-2337(1988).  
 RN [8]  
 RP REVISION TO 1184.  
 RX MEDLINE=89098346; PubMed=3211760;  
 RA Molyneux K., Dalglish R.;  
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
 RL Nucleic Acids Res. 16:11833-11833(1988).  
 RN [9]  
 RP SEQUENCE OF 1065-1466 FROM N.A.  
 RX MEDLINE=85087944; PubMed=6096827;  
 RA Loidi H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,  
 RA Rosebloom J., Myers J.C.;  
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
 RT procollagen.";  
 RL Nucleic Acids Res. 12:3383-3394(1984).  
 RN [10]  
 RP SEQUENCE OF 965-1200.  
 RX MEDLINE=81208139; PubMed=7016180;  
 RA Seyer J.M., Kang A.H.;

RT "Covalent structure of collagen: amino acid sequence of alpha  
 RT 1(III)-CBs from type III collagen of human liver.";  
 RL Biochemistry 20:2621-2627(1981).  
 RN [11]  
 RP SEQUENCE OF 1176-1466 FROM N.A.  
 RX MEDLINE=85157600; PubMed=2579949;  
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
 RA "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
 RT (III) collagen. Partial characterization of the 3' end region of the  
 RT gene.";  
 RL J. Biol. Chem. 260:4357-4363(1985).  
 RN [12]  
 RP SEQUENCE OF 1161-1200 FROM N.A.  
 RX MEDLINE=86187804; PubMed=3754462;  
 RA Miskulin M., Dalglish R., Kluwe-Beckerman B., Rennard S.I.,  
 RA Tolstochev P., Brantly M., Crystal R.G.;  
 RT "Human type III collagen gene expression is coordinately modulated  
 RT with the type I collagen genes during fibroblast growth.";  
 RL Biochemistry 25:1408-1413(1986).  
 RN [13]  
 RP SEQUENCE OF 1-170 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88303360; PubMed=3405773;  
 RA Toman D., Ricca G., de Crombrughe B.;  
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
 RT of human prepro alpha 1(III) collagen.";  
 RL Nucleic Acids Res. 16:7201-7201(1988).  
 RN [14]  
 RP SEQUENCE OF 1-176 FROM N.A.  
 RX MEDLINE=89378752; PubMed=777083;  
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
 RT "Cloning and analysis of the 5' portion of the human type-III  
 RT procollagen gene (COL3A1).";  
 RL Gene 78:255-265(1989).  
 RN [15]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuitaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [16]  
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RX MEDLINE=93293988; PubMed=8514865;  
 RA Tromp G., Wu Y., Prockop D.J., Madhatter S.L., Kleinert C.,  
 RA Early J.D., Zhang J., Noerregaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakkola P., Rymaszewski M., Pearce W.H., Yao J.S.T.,  
 RA Majamaa K., Smulens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michels V.V., Kaye M., Kuitaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
 RT in the triple-helical domain of type III procollagen are an  
 RT infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 RN [17]  
 RP VARIANT THR-698.  
 RX MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Kleinert C., Tromp G., Kuitaniemi H., Kontusaari S.,  
 RA Wu Y., Ganguly A., Prockop D.J.;  
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 RN [18]  
 RP VARIANT AORTIC ANEURYSM ARG-786.  
 RX MEDLINE=91056145; PubMed=2243125;  
 RA Kontusaari S., Tromp G., Kuitaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
 RT with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 RN [19]  
 RP VARIANT EDS-IV ARG-828.  
 RX MEDLINE=94016365; PubMed=8411057;  
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
 RT "The substitution of glycine 661 by arginine in type III collagen



RT produces mutant molecules with different thermal stabilities and  
 RT causes Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 30:690-693(1993).  
 RN [20]  
 RP VARIANT EDS-IV SER-957.  
 RX MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation in the type III procollagen gene that converts  
 the alpha 1 (III) chain of type III procollagen exposes an arginine  
 RT and causes Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:1349-1352(1989).  
 RN [21]  
 RP VARIANT EDS-IV VAL-960.  
 RX MEDLINE=95268429; PubMed=7749417;  
 RA Tromp G., de Paeppe A., Nuytink L., Madhatter S.L., Kuivaniemi H.;  
 RT "Substitution of valine for glycine 793 in type III procollagen in  
 RT Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 RN [22]  
 RP VARIANT EDS-IV GLU-1014.  
 RX MEDLINE=92316511; PubMed=1352273;  
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
 RA Pope F.M.;  
 RT "A single base mutation in the gene for type III collagen (COL3A1)  
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
 RT syndrome type IV. An unaffected family member is mosaic for the  
 RT mutation.";  
 RL Hum. Genet. 89:414-418(1992).  
 RN [23]  
 RP VARIANT EDS-IV ASP-1050.  
 RX MEDLINE=90037070; PubMed=2808425;  
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
 RT "Single base mutation in the type III procollagen gene that converts  
 RT the codon for glycine 883 to aspartate in a mild variant of  
 RT Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:19313-19317(1989).  
 RN [24]  
 RP VARIANT EDS-IV VAL-1077.  
 RX MEDLINE=91374480; PubMed=1895316;  
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,  
 RA Pope F.M.;  
 RT "Characterisation of a glycine to valine substitution at amino acid  
 RT position 910 of the triple helical region of type III collagen in a  
 RT patient with Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 28:458-463(1991).  
 RN [25]  
 RP VARIANT EDS-IV GLU-1173.  
 RX MEDLINE=93022543; PubMed=1357232;  
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
 RA Query Match 5.8%; Score 141.5; DB 1; Length 1466;  
 Best Local Similarity 27.6%; Pred. No. 0.055;  
 Matches 61; Conservative 13; Mismatches 74; Indels 73; Gaps 10;

ID CA18 RABIT STANDARD; PRT; 744 AA.  
 AC P14282;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).  
 GN COL8A1.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89380199; PubMed=2476437;  
 RA Yamaguchi N., Benvia P.D., van der Reest M., Ninomiya Y.;  
 RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs  
 RT demonstrate that type VIII collagen is a short chain collagen and  
 RT contains triple-helical and carboxyl-terminal non-triple-helical  
 RT domains similar to those of type X collagen.";  
 RL J. Biol. Chem. 264:16022-16029(1989).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT  
 CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.  
 CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION  
 CC WITH ALPHA 2(VIII) TYPE COLLAGENS.  
 CC -1- PTM: POLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPLETS ARE PRESENT  
 CC AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE  
 CC HIGH THERMAL STABILITY OF THIS REGION OF THIS REGION.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J05042; AAA1204.1; .  
 DR PIR; A34246; A34246.  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF01391; Collagen; 8.  
 DR PRINTS; PRO0007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Cell adhesion; Collagen; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 744 COLLAGEN ALPHA 1(VIII) CHAIN.  
 FT DOMAIN 29 117 NONHELICAL REGION (NC2).  
 FT DOMAIN 118 571 TRIPLE-HELICAL REGION.  
 FT DOMAIN 572 744 NONHELICAL REGION (NC1).  
 FT DOMAIN 609 744 C1Q.  
 SQ SEQUENCE 744 AA; 73358 MW; 2A8CE1FE8274E99 CRC64;  
 Query Match 5.8%; Score 141; DB 1; Length 744;  
 Best Local Similarity 23.6%; Pred. No. 0.029;  
 Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;



Db 273 KPEVTGPGGSLGKRPGEFPGPIGVPGVPGPLPGVGVKGGQGIPEQPGFPGG 332  
 QY 211 ATRGAEGLPPSPMPA-----GAAGLALLLVAGAGAMCRRR-----A 256  
 Db 333 KGGQGLPGPLPGVPGKPGFPGPKD-----RGIGVPGALGPRGSKGVAPGAWG 387  
 QY 257 KPESHPG-PSFGGSGSLG-----GGGGMGPR-----EAPGSLGIALKG--GAAD 304  
 Db 388 PPEPGPLPGIPGMPGPPGALGPPGKGGTGVPGPPGPKGPPGALQGGPPGFLGVBG 447  
 QY 305 P--PCPHYKVSQDYG-----PYIVQDGP-----QSPNNI 336  
 Db 448 PPGIRGLPGIPGKPGKAGKGLPGVPLGPGVPGKPGPIPDGQGLGPGI 499

RESULT 45  
 CA18 MOUSE STANDARD; PRT; 743 AA.  
 ID CA18 MOUSE  
 AC 000780; 09D2V4;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VIII) chain precursor.  
 GN COL8A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RX MEDLINE=92362626; PubMed=1499564;  
 RA Muragaki Y., Shiota C., Inoue M., Ooshima A., Olsen B.R.,  
 Nishimura Y.,  
 RT "Alpha 1(VIII)-collagen gene transcripts encode a short-chain  
 collagen polypeptide and are expressed by various epithelial,  
 RT endothelial and mesenchymal cells in newborn mouse tissues.",  
 RL Eur. J. Biochem. 207:895-902(1992).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Aichi U., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,  
 RA Saio T., Okazaki Y., Gojobori T., Bono H., Kaakawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Glass C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Stabile F., Suzuki R., Tomita M., Wagner L., Wasth T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Cantucci P., de Bernaldo M.F.,  
 RA Brownstein M.U., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamya W., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT  
 CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.  
 CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION  
 CC WITH ALPHA 2(VIII) TYPE COLLAGENS.  
 CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN CALVARIUM, EYE & SKIN OF  
 CC MEMBRANE CELLS, ALSO IN VARIOUS EPITHELIAL, ENDOTHELIAL AND  
 CC MESENCHYMAL CELLS.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; X66976; CA47387.1; --  
 DR EMBL; X66977; CA47387.1; JOINED.  
 DR EMBL; AK018742; BAB31383.1; --  
 DR PIR; S23779; S23779.  
 DR MGD; MGI:88463; Col8a1.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 7.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR SMART; SM00110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 RX Glycoprotein; Cell adhesion; Collagen; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 743  
 FT DOMAIN 29 117  
 FT DOMAIN 118 571  
 FT DOMAIN 572 743  
 FT DOMAIN 608 743  
 FT CONFLICT 6 6  
 FT CONFLICT 85 85  
 FT CONFLICT 109 109  
 FT CONFLICT 248 248  
 FT CONFLICT 313 313  
 FT CONFLICT 323 324  
 FT CONFLICT 361 361  
 FT CONFLICT 596 596  
 FT CONFLICT 717 719  
 SQ SEQUENCE 743 AA; 73454 MW; F584D58BD53897F4 CRC64;  
 Query Match 5.7%; Score 140.5; DB 1; Length 743;  
 Best Local Similarity 21.3%; Pred. No. 0.032;  
 Matches 101; Conservative 30; Mismatches 135; Indels 209; Gaps 21;  
 QY 14 GALLIGLVGLVGLSLSEPPYNNNSANKRFQAGGY-----LYQIGRRLDLCRRAPPG 69  
 Db 6 GPLQLLILPFTIS-----LNSV-RLIQAGAYYGKPLPPQIPPPQIPQYQPLG 55  
 QY 70 ---PH-----SSPNEYFKLY-----L 83  
 Db 56 QQVHMLKDGSLGSKEMPHMOYKPEPLPYMKIIPPVPMKGVVKKKGVPLAS 115  
 QY 84 VGGAGRRCE---APPAPMLLTCDRPDLDFTTIKFQYSNVLNGHERSHHYITA 139  
 Db 116 LRGGQGRGEPGPGPGP-----GGLPDHGMFG-----IK 146  
 QY 140 TSDST-----RGLGSLGG--GYCLTRGKYLIRVQSPRGAVPRKVPSEMPERRD 190  
 Db 147 GKPPQGVPGIGKGMGMGKPGAMGMPAK-----GEIPGKGIEMGIGFPG 196  
 QY 191 G-AAHSL---EPGKENTLPDPTSNATSRGAGPLPPPM----- 225  
 Db 197 GPPPGPLPGIGKGGGGLGQPGAKG-ERGPKPGPPGLQPGKGGKGMPLPOLKG 255  
 QY 226 -PAVAGAGGALLLVAGAGAMCRRARAPSSRHNG----- 265  
 Db 256 PPGHGGPFGVGLPGVKGVTGPPGQGLGKPGPGEGPGGLIGVPGVPPGMPGV 315  
 QY 266 -----PSFGGSGSLGCG--GGGKGP 287  
 Db 316 GKPGQDDIPQPGPPGKGGQGLPGPLPGVGLGKPGPPGKGGRGIGVGLGPR- 374  
 QY 288 AEPGELGIALRGGAADPPRCPHYKVSQDYGHPYIVQDGP-----PQSPF 334





ID CALB MOUSE STANDARD; PRT: 1804 AA.  
 AC 061245; 064047;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(XI) chain precursor.  
 GN COL1A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=96015067; PubMed=8530046;  
 RA Yoshida H., Inoguchi K., Khaleduzzaman M., Nimomiya Y.,  
 RA Andrikopoulos K., Ramirez F.,  
 RT "Coding sequence and alternative splicing of the mouse alpha 1(XI)  
 RT collagen gene (Col1a1).";  
 RL Genomics 28:337-340(1995).  
 [2]  
 RP SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=95163095; PubMed=7859283;  
 RA Li Y., Lacerda D.A., Warman M.L., Beier D.R., Yoshida H.,  
 RA Nimomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,  
 RA Ramirez F., Wardell B.B., Liffert G.D., Teuscher C., Woodward S.R.,  
 RA Taylor B.A., Seemiller R.E., Olsen B.R.,  
 RT "A fibrillar collagen gene, Col1a1, is essential for skeletal  
 RT morphogenesis";  
 RL Cell 80:423-430(1995).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)=1(II) (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DISEASE: DEFECTS IN COL1A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA,  
 CC AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS  
 CC CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES  
 CC AND TRACHEA.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
 CC  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: D38162; BAB07357.1; -  
 DR EMBL: S74574; AAB33439.1; -  
 DR MGI: 88446; Col1a1.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000085; Fib collagen\_C.  
 DR InterPro: IPR001731; Laminin G.  
 DR InterPro: IPR001230; Prenyl\_site.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PFO1391; Collagen; 16.  
 DR Pfam: PFO1410; COLFT; 1.  
 DR Pfam: PFO2210; TSPN; 1.  
 DR ProDom: PD000007; Collagen; 1.  
 DR ProDom: PD002078; Fib collagen\_C; 1.  
 DR SMART: SM00038; COLFT; 1.  
 DR SMART: SM00282; LamG; 1.  
 DR SMART: SM00210; TSPN; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KM Glycoprotein; Collagen; Signal; Alternative splicing;  
 KM Disease mutation.  
 FT STGNL 1 35  
 FT PROPEP 36 511  
 FT CHAIN 512 1561  
 FT PROPEP 1562 1804  
 FT DOMAIN 37 417  
 FT DOMAIN 418 506  
 FT DOMAIN 507 509  
 FT DOMAIN 510 527  
 FT DOMAIN 528 1540  
 FT DOMAIN 1541 1561  
 FT CARBOHYD 1638 1638  
 FT SITE 610 610  
 FT SITE 1450 1450  
 FT VARSPLIC 329 413  
 FT VARIANT 182 195  
 FT VARIANT 196 1804  
 SQ SEQUENCE 1804 AA; 180963 MW; FE2DB9DED1E4219A CRC64;  
 Query Match 5.7%; Score 139; DB 1; Length 1804;  
 Best Local Similarity 25.7%; Pred. No. 0.099;  
 Matches 105; Conservative 17; Mismatches 116; Indels 170; Gaps 24;  
 QY 2 GPP-HSGPGGVRGALLLGLVGL--VSGLSLEPVYWNANRFGAEGGVLYPQIGDL 58  
 DB 440 GPPGPAPGAPG-----LWGPPLQGPSPGLPGDP-----GDR- 469  
 QY 59 DLICPRAPPGPHSSPNRYKYLVGAQGRCEAPAPNLLTCDRDLDRF----- 113  
 DB 470 -----GPPGRGGLP-----GADG-LPGPPGWM-----LPRRYGDDG 502  
 QY 114 ---TIKQEVSPNLMGHEFRSHDYIATSDTRGLESLOGCVLTGKMTLRYG-- 168  
 DB 503 KGPFIHQEAQAQ-----ALIQARIALGPPGMLTG-----PPGPV 541  
 QY 169 -----QSPRGAVPRKPYSE-----MPMER-----DRG-- 191  
 DB 542 GGPESTGAKESGDPGPGRGVQGPFGPKGCKGRPGADGGRGMPGSGSGKRGFD 601  
 QY 192 -----AAHSLEPGKENTPGDPTSNATSRGAEGPLPPSMBAVAGAAAGLALLLGA 243  
 DB 602 GLPEPLPDKHKRGRGPGPGLPGDDG-WRGEDGIGPGLGEGAGPRG-----LLGPR 655  
 QY 244 GA-----GAACRRAPRAAPSESRRNG-PGSGRGSGSLG--GGGGMGRREAPGE 292  
 DB 656 GTPGPPOGPFGIGIDGPGQKGNMGPGEPGPGQGNPESQGLPAPQGTGP-----PGE 711  
 QY 293 LGIALRGG-----GAADPPCPHYEKVSGDYGHVYIVDGP--PQSP 333  
 DB 712 KPGQKRGGLAGLRGADGPPGHPKRGSGSGKG-----ALGPPGPQGP 753  
 RESULT 49  
 ID CAL2 MOUSE STANDARD; PRT: 1459 AA.  
 AC P28481;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].  
 GN COL2A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91358489; PubMed=1885613;  
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.,  
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
 RT structure, and alternative splicing";  
 RL J. Biol. Chem. 266:16862-16869(1991).

RN SEQUENCE OF 1455-1459 FROM N.A.  
 RP MEDLINE=91274355; PubMed=2054384;  
 RX Metaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse type I, II, III and IX  
 collagen mRNAs."; Acta 1089:241-243(1991).  
 RL Bioclim. Biophys.  
 CC -1- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.  
 CC -1- PTM: POLYMERIZES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 1 VWFPC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 CC EMBL; M65161; AAA68100.1; -  
 CC EMBL; X57982; CAA41047.1; -  
 CC MGD; MG1:86452; Col2a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; FIB\_collagen\_C.  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF00093; vwc; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF01410; COLFI; 1.  
 DR ProDom; PD000007; Collagen; 3.  
 DR ProDom; PD002078; FIB\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; VWFPC; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycopolymers; Collagen; Cartilage; Signal; Alternative splicing.  
 FT STGM1 1 25  
 FT PROPEP 26 153  
 FT CHAIN 154 1213  
 FT PROPEP 1214 1459  
 FT DOMAIN 32 89  
 FT DOMAIN 173 1186  
 FT VARSPLIC 29 29  
 FT VARSPLIC 30 98  
 SQ SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E7F2 CRC64;  
 Query Match 5.7%; Score 138.5; DB 1; Length 1459;  
 Best Local Similarity 30.1%; Pred. No. 0.086; Indels 47; Gaps 9;  
 Matches 56; Conservative 13; Mismatches 75;  
 QY 173 GGAIVPRKPVSEMERDRGAHSLPEKENT-----PGD--PTSNTSRGAEPLPPPS 224  
 Db 470 GGAAPRIGPGRGAPGRNGF-----PGDGLAGPKAPGRGRSGLAGRPGANGDPGRFC 524  
 QY 225 MEAVAGAGAGIALLLLVAGAGGMCWRRRAKPSRSRHGP-----GSFRGSLGLGGG 280  
 Db 525 EPGLPGRG-----LGRPGDAGPQGVPSGAGEDGRGPGPQARQPGVMGPPGP 579  
 QY 281 GGMGPRAEPGLGIA-----LRG-----GGAADPPFCFHYEKVSGDYGHPIVYVDGP 329  
 Db 580 KGANGEPKAGKGLAGPGLRGIPGIDGRTGAAGCPGPGSPAGREDEQAP-----GP 633  
 QY 330 -----PQSP 334  
 Db 634 SGFOGLPGRPPGP 646  
 RESULT 50  
 CA2B\_HUMAN

ID CA2B\_HUMAN STANDARD; PRT; 1736 AA.  
 AC P13942; Q13273; Q13272; Q07751; Q99866; Q9UIP9;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 2(XI) chain precursor.  
 GN COL11A2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96037717; PubMed=7559422;  
 RA Vuorio E.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,  
 RA Ala-Kokko L.;  
 RT "The human COL11A2 gene structure indicates that the gene has not  
 RT evolved with the genes for the major fibrillar collagens."; J  
 RT J. Biol. Chem. 270:22673-22881(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tubby B.;  
 RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 59-807 FROM N.A.  
 RC TISSUE=Cartilage; PubMed=8325374;  
 RX MEDLINE=93314796;  
 RA Zhidkova N.I., Brewton R.G., Mayne R.;  
 RT "Molecular cloning of PAPP (proline/arginine-rich protein) from human  
 RT cartilage and subsequent demonstration that PAPP is a fragment of the  
 RT NH2-terminal domain of the collagen alpha 2(XI) chain."; J  
 RT FEBS Lett. 326:25-28(1993).  
 RN [4]  
 RP SEQUENCE OF 730-1690 FROM N.A.  
 RX MEDLINE=89340485; PubMed=2760050;  
 RA Kimura T., Cheah K.S.B., Chan S.D.H., Lui V.C.H., Mattei M.-G.,  
 RA van der Rest M., Ouo K., Solomon E., Ninomiya Y., Olsen B.R.;  
 RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of  
 RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen  
 RT with differences in genomic organization."; J  
 RT J. Biol. Chem. 264:13910-13916(1989).  
 RN [5]  
 RP SEQUENCE OF 1-537 FROM N.A.  
 RX MEDLINE=96435918; PubMed=8838804;  
 RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;  
 RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding  
 RT information, identification of the promoter sequence, and precise  
 RT localization within the major histocompatibility complex reveal  
 RT overlap with the K85 gene."; J  
 RT Genomics 32:401-412(1996).  
 RN [6]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=95238468; PubMed=7721876;  
 RA Zhidkova N.I., Justice S.K., Mayne R.;  
 RT "Alternative mRNA processing occurs in the variable region of the  
 RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains."; J  
 RT J. Biol. Chem. 270:9486-9493(1995).  
 RN [7]  
 RP DISEASE.  
 RX PubMed=10677296;  
 RA Meloniemi M., Brunner H.G., Manouvrier S., Hennekam R.,  
 RA Superfi-Furga A., Kaerleirinen H., Pauli R.M., van Biesen T.,  
 RA Warman M.L., Bonaventura J., Mitty P., Ala-Kokko L.;  
 RT "Autosomal recessive disorder otospondylosynostosis/dysplasia is  
 RT associated with loss-of-function mutations in the COL11A2 gene."; J  
 RT Am. J. Hum. Genet. 66:368-377(2000).  
 RN [8]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97255959; PubMed=9101290;  
 RA Kuitvanen H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";

RL Hum. Mutat. 9:300-315(1997).  
 RN [9]  
 RP VARIANT OSMED ARG-661.  
 RX MEDLINE=95163096; PubMed=7859284;  
 RA Vilkku M., Marttinen E.C.M., Lai V.C.H., Zhidkova N.I., Tiller G.E.,  
 RA Goldring M.B., van Beersum S.E.C., de Maal Malfijt M.C.,  
 RA van den Hoogen F.H.J., Ropers H.-H., Mayne R., Cheah K.S.E.,  
 RA Olsen B.R., Warman M.L., Brunner H.G.,  
 RT "Autosomal dominant and recessive osteochondrodysplasias associated  
 RT with the COL11A2 locus".  
 RL Cell 80:431-437(1995).  
 RN [10]  
 RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.  
 RX PubMed=9585596;  
 RA Koga H., Sakou T., Takeuchi E., Hayashi K., Numasawa T., Harata S.,  
 RA Yone K., Matsunaga S., Ohterud B., Inoue I., Leppert M.,  
 RT "Genetic mapping of ossification of the posterior longitudinal  
 RT ligament of the spine".  
 RL Am. J. Hum. Genet. 62:1460-1467(1998).  
 RN [11]  
 RP VARIANT W2S GLU-955.  
 RX PubMed=9805126;  
 RA Philajmaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,  
 RA Gleason A., Wiesbauer P., Spranger J., Ala-Kokko L.,  
 RT "Heterozygous glycine substitution in the COL11A2 gene in the original  
 RT patient with the Weissenbacher-Zweymueller syndrome demonstrates its  
 RT identity with heterozygous OSMED (nonocular Stickler syndrome)".  
 RL Am. J. Med. Genet. 80:115-120(1998).  
 RN [12]  
 RP VARIANT STL3 940-GLY-PRO-948 DEL.  
 RX PubMed=9506662;  
 RA Sirko-Osada D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,  
 RA Robin N.H.,  
 RT "Stickler syndrome without eye involvement is caused by mutations in  
 RT COL11A1, the gene encoding the alpha-2(XI) chain of type XI  
 RT collagen".  
 RL J. Pediatr. 132:368-371(1998).  
 RN [13]  
 RP VARIANTS DFNA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.  
 RX MEDLINE=20047768; PubMed=10581026.  
 RA McGlite W.T., Praad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,  
 RA Shergel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,  
 RA King M.-C., Brunner H.G., Cremers C.W.R.J., Takano S.W.,  
 RA Arima M., Mayne R., Prockop D.J., Van Camp G., Smith R.T.H.,  
 RT "Mutations in COL11A2 cause non-syndromic hearing loss (DFNA13)".  
 RL Nat. Genet. 23:413-419(1999).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILGENESIS BY  
 CC -1- CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 3(XI) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)=1(II).  
 CC -1- ALTERNATIVE PRODUCTS: 8 isoforms; 1 (shown here), 2, 3, 4, 5, 6, 7  
 CC and 8; may be produced by alternative splicing. They lack exons 6,  
 CC 7 or 8 or a combination of these exons.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH  
 CC PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING  
 CC EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE  
 CC CARILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT  
 CC AMOUNTS.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of Stickler syndrome  
 CC type 3 (STL3). It is an autosomal dominant disorder characterized  
 CC by oro-facial, auditory and skeletal manifestations, such as  
 CC midfacial hypoplasia, cleft palate, osteoarthritis, and  
 CC sensorineural hearing loss. Differently from Stickler syndrome  
 CC type 1 and 2, no ocular involvement is observed. This disorder is  
 CC also referred to as Stickler-like syndrome or non-ocular Stickler  
 CC syndrome.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal recessive  
 CC osteochondrodysplasia (OSMED), a skeletal dysplasia  
 CC accompanied by severe hearing loss. The phenotype overlaps that of

CC autosomal dominant skeletal disorders (Stickler and Marshall  
 CC syndromes) but can be distinguished by disproportionately short  
 CC limbs and lack of ocular involvement.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of Weissenbacher-  
 CC Zweymueller syndrome (W2S), an autosomal dominant disorder allelic  
 CC with STL3 and OSMED. It is also referred to as heterozygous OSMED.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal dominant  
 CC nonsyndromic sensorineural deafness type 13 (DFNA13). Affected  
 CC individuals experience progressive hearing loss beginning in the  
 CC second to fourth decades, eventually making use of amplification  
 CC mandatory.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- DATABASE: NAME=Hereditary hearing loss homepage;  
 CC WWW="http://www.uia.ac.be/dna1ab/hhh/hhgenes.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U32169; AAC50214.1; -  
 DR EMBL; U32169; AAC50213.1; -  
 DR EMBL; U32169; AAC50215.1; -  
 DR EMBL; AA031226; CAA20240.1; -  
 DR EMBL; L18987; AAA55498.1; -  
 DR EMBL; J04974; AAA52034.1; -  
 DR EMBL; U41069; AAC17464.1; -  
 DR EMBL; U41069; AAC17464.1; JOINED.  
 DR EMBL; U41066; AAC17464.1; JOINED.  
 DR EMBL; U41067; AAC17464.1; JOINED.  
 DR PIR; A32645; A32645.  
 DR Genew; HGNC:2187; COL11A2.  
 DR MIM; 120290; -  
 DR MIM; 184840; -  
 DR MIM; 277610; -  
 DR MIM; 601868; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib.collagen\_C.  
 DR InterPro; IPR001791; Lamlnh\_G.  
 DR InterPro; IPR001230; Prenyl\_site.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF01410; COLF1; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR ProDom; PD002078; Fib.collagen\_C; 1.  
 DR SMART; SM00038; COLF1; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.

Query Match 5.7%; Score 138.5; DB 1; Length 1736;  
 Best Local Similarity 31.7%; Pred. No. 0.1; indels 49; Gaps 14;  
 Matches 64; Conservative 18; Mismatches

QY 147 GLSLGGVCLTGMVYLKLVGSPRGAVPRK--VVSSEWVERDRGAHSLRPGKENTP 204  
 DB 1327 GSEGRGG--KGA-----GDFGATGAPKATGVPR-----AGPAGKPGDGLR 1368  
 QY 205 GDFPS-----NATSR-GAEGFLPPSPMPAYAGAGGLAL---LIGVAGAGAMCMR 252  
 DB 1369 GLPFSVQQGRPATGQAGPPGVPGRPGRLGRLGDAGAKGKHPDLIGLPGG----- 1423  
 QY 253 RRKAKESESHNRP-GSFGGSGSLG--GGGGGGMGRREARPGELGIA-LRGGGAADRPGR 309  
 DB 1424 -EGEKGRGLRPPQSGPGKGWGI-PGASGPIGR-GRRPLGRPGKAGAKATGPGGR 1481  
 QY 310 HYEK-VSGDYGHPVYIVVDGPP 330

Wed Feb 12 09:30:11 2003

us-10-021-121-2.rsp

Page 41

Db 1482 KGEKVGPPGHP-----GPP 1497

Search completed: February 11, 2003, 12:03:44  
Job time : 32.4591 secs



Db 165 PGAPGPPGPAEKAGKAMKRDATPSGPGPPGVKGAAG-----LQGPQAPG----- 214  
QY 253 RRAKQSSSHHP-----GSFGGSLGLGQGGGMPRAEPFELGI-----ALRGGG 301  
Db 215 ----KQATGTTPGPQGEKSGKGGGLIGPKGFTG---TGKXKDJGLPSKSKDRGKGDPA 267  
QY 302 AADPPCPHTEKXSGDYGHVYIVQDGP 330  
Db 268 GVMGP--PGAQSGKSDPGRP-----GPP 288

RESULT 50  
US-09-978-192A-614  
Sequence 614, Application US/09978192A  
Patent No. US2002017753A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnovers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Feirata, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austen L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavlin, Iyar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James J.  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Thomas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
Acids Encoding the Same  
FILE REFERENCE: P2630PIC9  
CURRENT APPLICATION NUMBER: US/09/978,192A  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918565  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082566  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797



;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082796  
;; PRIOR FILING DATE: 1998-04-23  
;; PRIOR APPLICATION NUMBER: 60/083336  
;; PRIOR FILING DATE: 1998-04-27  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083392  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083495  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083496  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083499  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083545  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083554  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083558  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083500  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083742  
;; PRIOR FILING DATE: 1998-04-30  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084441  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084637  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084598  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084600  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084627  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085689  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 5.2%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.9%; Pred. No. 0.062;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 PGKENTLPDEPTSNATS-----RGABGP.LPPSPMVAAGAGLALILLGVAGGAMCMR 252  
Db 165 PGADPPGPPEAKGAKGAMGRDGTGSPGQPPGVKGEAG-----LQPGAPG----- 214  
QY 253 RRAKPESESRHGP-----GSFGRGSLGLGGGGGMPREAPPELIG-----ALRGGG 301  
Db 215 ----KQATGTGPQGEGSKGDSGLIGPKETG---TKGKQDLGLPSKGRGKMGDA 267  
QY 302 AADPPCFPHYEXVSGDYGHAPVYIVDGP 330  
Db 268 GVMGP--PGAQSGKDPGRP-----GPP 288

Search completed: February 11, 2003, 12:13:11  
Job time : 25.4591 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 11:47:56 ; Search time 33.7673 Seconds

(Without alignments)  
2776.398 Million cell updates/sec

Title: US-10-021-121-2

Perfect score: 2450

Sequence: 1 MGPPHSGPGVAVGALLLG.....TTLRGRASVEAAGQHPL 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 50 summaries

```

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORNITHINE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_XYLOPHAGUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1550	63.3	285	4	Q8TBH7
2	914	37.3	331	13	Q90Z31
3	910	37.1	327	13	Q9PT69
4	620.5	25.3	333	13	Q9PTU4
5	596.5	24.8	341	13	Q90Z33
6	596.5	24.3	334	13	Q90Z32
7	331	13.5	205	13	Q9W6H9
8	214.5	8.8	237	5	Q9U3M2
9	198.5	8.1	279	5	Q9U3M2
10	192	7.8	652	5	Q9V4B1
11	178.5	7.3	202	13	Q98T21
12	167	6.8	205	11	Q9D7K8
13	163.5	6.7	206	11	Q9CZS8
14	155.5	6.3	675	6	Q9N178
15	155	6.3	229	13	Q93431
16	153.5	6.3	1691	11	Q9BSQ2

17	151.5	6.2	1447	13	Q91B91	Q91B91 xenopus lae
18	149	6.1	325	5	Q17036	Q17036 caenorhabd
19	149	6.1	569	5	Q17208	Q17208 bombyx mori
20	148.5	6.1	316	5	Q19111	Q19111 caenorhabd
21	146.5	6.0	1140	11	Q61434	Q61434 mus musculu
22	146.5	6.0	1491	13	Q91718	Q91718 xenopus lae
23	146.5	6.0	1774	11	Q62001	Q62001 mus musculu
24	146	6.0	675	13	Q90800	Q90800 gallus gall
25	146	6.0	1669	11	Q9QZS0	Q9QZS0 mus musculu
26	145.5	5.9	305	5	Q17805	Q17805 caenorhabd
27	145.5	5.9	308	5	Q94620	Q94620 meloidogyne
28	145	5.9	680	11	Q9DD02	Q9DD02 mus musculu
29	145	5.9	1745	5	Q9NZ06	Q9NZ06 homo sapien
30	144.5	5.9	309	5	Q25456	Q25456 meloidogyne
31	144.5	5.9	1347	4	Q96Q53	Q96Q53 homo sapien
32	144.5	5.9	1420	13	Q90W37	Q90W37 gallus gall
33	143.5	5.9	775	16	Q9F342	Q9F342 streptomyce
34	143	5.8	1747	5	Q26640	Q26640 strongyloce
35	142	5.8	614	5	Q76368	Q76368 caenorhabd
36	142	5.8	744	4	Q96D07	Q96D07 homo sapien
37	142	5.8	1835	13	Q91NU4	Q91NU4 gallus gall
38	141	5.8	284	5	Q25582	Q25582 teladorsagi
39	141	5.8	284	5	Q25581	Q25581 teladorsagi
40	141	5.8	319	5	Q17038	Q17038 caenorhabd
41	141	5.8	809	13	Q93485	Q93485 oncorhynch
42	141	5.8	1449	13	Q910C0	Q910C0 oncorhynch
43	140.5	5.7	219	13	Q90YC5	Q90YC5 brachydanio
44	140.5	5.7	305	5	P91274	P91274 caenorhabd
45	140	5.7	1378	5	Q9XU59	Q9XU59 caenorhabd
46	140	5.7	1378	5	Q9Y1B4	Q9Y1B4 halictis di
47	139.5	5.7	1450	13	Q9Y1B4	Q9Y1B4 cynops pyrr
48	139.5	5.7	1690	4	Q9UMG6	Q9UMG6 homo sapien
49	139.5	5.7	1691	4	Q9Y414	Q9Y414 homo sapien
50	139	5.7	705	4	Q8TBZ5	Q8TBZ5 homo sapien

## ALIGNMENTS

RESULT 1	Q8TBH7	PRELIMINARY:	PRT:	285 AA.
AC	Q8TBH7			
DT	01-JUN-2002 (TREMBLER, 21, Created)			
DT	01-JUN-2002 (TREMBLER, 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLER, 21, Last annotation update)			
DE	Similar to ephrin B3 (fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RA	Strasbourg R.			
RL	Submitted (F88-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC022499; AAH22499.1; -.			
FT	NON TER			
SQ	SEQUENCE 285 AA; 30159 MW; 08D8ACE7142469B2 CRC64;			
Query Match	63.3%; Score 1550; DB 4; Length 285;			
Best local similarity	100.0%; Pred. No. 6,2e-114;			
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	56 DRDLICRRAPPQPHSSPNTFFKYLIVGAGQRCCEAPPANLLITCDPDDLDLFTI 115			
DB	1 DRDLICRRAPPQPHSSPNTFFKYLIVGAGQRCCEAPPANLLITCDPDDLDLFTI 60			
QY	116 KFOEVSPTLWGERSHRDYIIATSDGTRRGLESIGGVCLTGKTVLIRVGSPPGGA 175			
DB	61 KFOEVSPTLWGERSHRDYIIATSDGTRRGLESIGGVCLTGKTVLIRVGSPPGGA 120			
QY	176 VPRKVSFEMPERDRGAHSLPEKKNLPQDPSTNATSRGAEGPLPPSPMPAVAGAAGL 235			

```

Db 121 VRRKPVSEKPMERDRGAASLSEPGKENTLPGDPTSNATSGAGCPLPPSPMAVAAGAAGL 180
QY 236 ALLILGVAAGAGAMCWRRRRAKPSRRHAPGSGSLGIGGGGCMGPRRAEAGELGI 295
Db 181 ALLILGVAAGAGAMCWRRRRAKPSRRHAPGSGSLGIGGGGCMGPRRAEAGELGI 240
QY 296 ALRGGGAADPPFCPHYEKVSGDYGHVYVQDGPQSPSPNNIY 338
Db 241 ALRGGGAADPPFCPHYEKVSGDYGHVYVQDGPQSPSPNNIY 283

RESULT 2
Q90231 PRELIMINARY; PRT; 331 AA.
ID 090231
AC 090231
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Ephrin B3.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21290827; PubMed=11397014;
RA Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,
RA Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;
RT "Morphogenesis of prechordal plate and notochord requires intact
RT eph/ephrin b signaling.";
RL Dev. Biol. 234:470-482(2001).
DR EMBL; AF375227; AAK64277.1; -.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR Prodom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; UNKNOWN_1.
SQ SEQUENCE 331 AA; 35638 MM; 6A5EACD509A09818 CRC64;

Query Match 37.3%; Score 914; DB 13; Length 331;
Best Local Similarity 54.5%; Pred. No. 6.4e-64;
Matches 188; Conservative 43; Mismatches 74; Indels 40; Gaps 10;

QY 10 GYVAGALLLGLVGLVSGLSLEPVYNSANKRFQAEQGVLYVPOIGRDLILCPRAKPRG 69
Db 9 GIGILLIFVLDLG-ITATNMEPIYNSLNKRFSDKGYLYVPOIGRDLILCPSSDPG 67
QY 70 PHSSPVYEFKLYLVGG-AQGRRCAPAPNLLTCDRPDLRLFTIKFOEYSPNLMGHE 128
Db 68 PRAPADYEVYKLYLVSSRQADRCVETGAPNLLTCDKNSDMFTIKFOEYSPNLMGHE 127
QY 129 FRSHDYIYIANSYDGTREGLESIQGVCLTRGMKYLRYGSGPRG-GAVPRKPVSEMPNE 187
Db 128 FKTNHDFYFIATSDGTROGLESNRGVCAIQGMKYLKYGSGPRG-PASPRPS- 182
QY 188 RDRGAHSLSEPGKENTLPGDPTSNAT-----SRGAEPLPPSPMAVAAGAAGLALL 239
Db 183 -----AGRINPNPCTGNSTHPQIPPRGSGSENPPLASNTAVIAGAAGSAPFL 232
QY 240 LGVAGAGAMCWRRRRAKPSRRHAPGSGF-----RGSLGIGGGGMPREAPGEL 293
Db 233 L-TAVIACVCCYRRRAKSHSEHHP-PLGISLTSPPKCGGCGVGGANNNG---SEPSDI 287
QY 294 GILRGGGAADPPFCPHYEKVSGDYGHVYVQDGPQSPSPNNIY 338
Db 288 IIPLR---TSDAYCPHYEKVSGDYGHVYVQDGPQSPSPNNIY 329

RESULT 3
Q9PT69 PRELIMINARY; PRT; 327 AA.
ID 09PT69
AC 09PT69;

```

```

DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Ephrin-B3 precursor.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRIO;
RX MEDLINE=20099673; PubMed=10633856;
RA Heibling P.M., Saulnier D.M.E., Robinson V., Christiansen J.H.,
RA Wilkinson D.G., Brandt A.W.;
RT "Comparative analysis of embryonic gene expression defines potential
RT interaction sites for Xenopus EphA4 receptors with ephrin-B ligands.";
RL Dev. Dyn. 216:361-373(1999).
DR EMBL; AJ236866; CAB6511.1; -.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR Prodom; PD002533; Ephrin; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
DR PROSITE; PS01299; EPHRIN; 1.
KM Signal.
FT SIGNAL.
SQ SEQUENCE 327 AA; 35913 MM; 4BB0FA394C22DCD CRC64;

Query Match 37.1%; Score 910; DB 13; Length 327;
Best Local Similarity 60.1%; Pred. No. 1.3e-63;
Matches 190; Conservative 30; Mismatches 82; Indels 14; Gaps 8;

QY 25 VSGLSLEPVYNSANKRFQAEQGVLYVPOIGRDLILCPRAKPRGSHSSPNTYFYLYY 84
Db 22 ISALSLDPIYNSNKRFEDETEGYLYVPOIGRDLILCPRSSEPGFSSPYEYLYLYY 81
QY 85 GAGAG-RRCAPAPNLLTCDRPDLRLFTIKFOEYSPNLMGHEFRSHDYIYIANSYD 143
Db 82 GYKEMSSGSLTRPMLLTCDRPSQDLRLFTIKFOEYSPNLMGHEQSDRYIYIANSYD 141
QY 144 TREGLSLOQGVCLTRGMKYLRYVGS PRGAVPRKPVSEMPERDRGAASL-EPGKEN 202
Db 142 TMGIETLQGVCEYCKMKTLYKVGSPNGATPRRPS--AGKDSGISPVNPDIEN 198
QY 203 LPQDPTSNATSRGAEPLPPSPMAVAAGAAGLALLLGVAGAGAMCWRRRRAKPSER 262
Db 199 V-GETSNATKTGEMGFLPSHVLVAGAAGLALLL-VFGVGVVCHRRQAKASDTR 256
QY 263 HPGGSGFRGSGSLGIGGGGMPREAPGELGILRGGGAADPPFCPHYEKVSGDYGHV 322
Db 257 HP-PLGISLTSPPKCGGNN-----NGHEPSDIIMPLRSEAG-AFCPHYEKVSGDYGHV 309
QY 323 YIVODGPQSPSPNNIY 338
Db 310 YIVODMASOSPANNIY 325

RESULT 4
Q9PU04 PRELIMINARY; PRT; 333 AA.
ID 09PU04
AC 09PU04;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Ephrin-B2 precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Menzel P., Pasquale E.B.,  
 RT "Coding sequence of chicken ephrin-B2,"  
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF180729; AAD33948.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; Ephrin; 1.  
 KM Signal.  
 FT SIGNAL.  
 FT CHAIN  
 SQ SEQUENCE 333 AA; 36761 MW; 4C298CB211B7783 CRC64;

Query Match 25.3%; Score 620.5; DB 13; Length 333;  
 Best Local Similarity 40.2%; Pred. No. 7.2e-41;  
 Matches 133; Conservative 52; Mismatches 125; Indels 21; Gaps 6;

QY 14 GALLLGLVLSGLSPVYNSANKRFOAEAGVYLTPQIGRDLDCPRARPPGPHSS 73  
 DB 16 GALVLTALAKSIVLDPITYNSSNPKFLPGQLVLYPQIGRDLDCPRV---DSKTV 72  
 QY 74 PNYEYKLYLVGAQGRCEAPRNLLTCDRDLRFITKQEVSPNLWGHEFRSH 133  
 DB 73 GQYEVYKVMYDKQADSCARKNTPPLNCAKPDQVKTIFKQEPSPNLWGHEFRSH 132  
 QY 134 DYIATSDGTREGESLQGVCLTRGKVLRLVQSGPRGAVPRKVSSEMPER---- 188  
 DB 133 DYIATSDGTREGESLQGVCLTRGKVLRLVQSGPRGAVPRKVSSEMPER---- 189  
 QY 189 -DRGAHSLBEGKNTLPDPTSNATSRGAEGLPPPMVAVAGAGLALLGLVAGAG 247  
 DB 190 GTNGKSTSPFVVDHSGSTG--SKAGHSILGSEVALFAGIASCIPIVITLV 247  
 QY 248 AMCWRRRAKSESRHPGSGFSGSGSLGGGGGMPREAPBELIALRGGAAPPF 307  
 DB 248 LLLKXRRHRHRSQHTTLLSLSTLAPRKSGNN---NSGSEPDIIIPLR---TADSVF 300  
 QY 308 CPHEKVSQDYGHPIYIVQDGPSPSPNITY 338  
 DB 301 CPHEKVSQDYGHPIYIVQDGPSPSPNITY 331

RESULT 5  
 Q90233 PRELIMINARY; PRT; 341 AA.

AC Q90233;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Ephrin B1.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 CX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21290827; PubMed=11397014;  
 RA Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,  
 RA Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.,  
 RT "Morphogenesis of prechordal plate and notochord requires intact  
 RT eph/ephrin b signaling,"  
 RL Dev. Biol. 234:470-482(2001).  
 DR EMBL, AF375224; AAK64274.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; Ephrin; 1.  
 SQ SEQUENCE 341 AA; 37849 MW; CB922F20E0D93B94 CRC64;

Query Match 24.8%; Score 607; DB 13; Length 341;  
 Best Local Similarity 41.5%; Pred. No. 8.5e-40;  
 Matches 139; Conservative 49; Mismatches 109; Indels 38; Gaps 11;

QY 24 LVSGLSLEPVYNSANKRFOAEAGVYLTPQIGRDLDCPRARPPGPHSSPNYEFKYL 83  
 DB 23 LPAKSLSEPVYNSANKRFOAEAGVYLTPQIGRDLDCPRARPPGPHSSPNYEFKYL 77  
 QY 84 VQAGAGRCRAPRNLLTCDRDLRFITKQEVSPNLWGHEFRSHHDYIATSDG 143  
 DB 78 VKKEQAESCTLLDPRNVLVTKPKPKDKITFKQEPSPNLWGHEFRSHHDYIATSDG 137  
 QY 144 TRGLESLOGVCLTRGKVLRLVQSGPRG----GAVPRKVSSEMPERDGAHSLP 198  
 DB 138 TQEGLENRGVCGSTRKMKIIMKVGDDNAPDPLDPRDYDNEIKDPTSPSKTER 197  
 QY 199 GKEN-----LRDPTSNATSR--GAEGLPPSPMVAAGAG-GALLLGLVAGAG 247  
 DB 198 GENEVDGNGSMPPKDTNQNNSPGSVGLP--GSKPALPAAGAGCVITLLITLV 255  
 QY 248 AMCWRRRAKSESRHPGSGFSGGSL---GLGGGGGMPREAPBELIALRGGA 303  
 DB 256 LLLKLRKTR---KHSQ---RGSTLSLSTLAPRGAAGASEPDIIIPLR---TT 304  
 QY 304 DPPCPHEKVSQDYGHPIYIVQDGPSPSPNITY 338  
 DB 305 ENNYCPHEKVSQDYGHPIYIVQDGPSPSPNITY 339

RESULT 6  
 Q90232 PRELIMINARY; PRT; 334 AA.

AC Q90232;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Ephrin B2b.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 CX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21290827; PubMed=11397014;  
 RA Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,  
 RA Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.,  
 RT "Morphogenesis of prechordal plate and notochord requires intact  
 RT eph/ephrin b signaling,"  
 RL Dev. Biol. 234:470-482(2001).  
 DR EMBL, AF375226; AAK64276.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR ProDom; PD002533; Ephrin; 1.  
 SQ SEQUENCE 334 AA; 36998 MW; 341497E0FE9473BD CRC64;

Query Match 24.3%; Score 596.5; DB 13; Length 334;  
 Best Local Similarity 40.6%; Pred. No. 5.5e-39;  
 Matches 132; Conservative 45; Mismatches 113; Indels 35; Gaps 8;

QY 30 LEPVYNSANKRFOAEAGVYLTPQIGRDLDCPRARPPGPHSSPNYEFKYLTVGAGG 89  
 DB 27 LSIYVNTSNTKTFVGRGVLYLPQIGDKMDIVCPRIK-PGSEQGNIRHFRVLYPKQL 85  
 QY 90 RRCRAPRNLLTCDRDLRFITKQEVSPNLWGHEFRSHHDYIATSDGTREGLE 149  
 DB 86 ETCVYTKSDMLLNDKDPQDVKTFFKQEPSPNLWGHEFRSHHDYIATSDGTREGLE 145  
 QY 150 SLQGVCLTRGKVLRLVQSGPRGAV---PRKVSSEMPERDGAHSLBEGKENDPG 205  
 DB 146 NHGAGVCSKSKVLRLVQSGPDSFSAKNHPTNPPRYPENKQNTF---SKENDVS 200  
 QY 206 --DPTSNATSRGAEGLPPSPMVAAGAGLALLGLVAGAGMCM-----R 253  
 DB 201 QIDSKMNGSGGKSG-----BSVSGAGSDVALFAGV--ASGAVIPILIIALVALH 250

```

QY 254 RRAKPSERHPPGSGFGRGSLGLGGGGMGPAREPGLIARCGGAADPEFCHEK 313
DB 251 RRQKXSAQSGGULPNTLPKRGSGAGSGNNNSPDIIFIRKSGM---YCPHEK 307
QY 314 VSGDYGHPIYIVDGPPOSPPNIY 338
DB 308 VSGDYGHPIYIVDGPPOSPPNIY 332

RESULT 7
Q9W6H9 PRELIMINARY; PRT; 205 AA.
AC Q9W6H9;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Epirin-B2 (fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodidae; Xenopus.
NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97411149; PubMed=9259557;
RA Smith A., Robinson V., Patel K., Wilkinson D.G.;
RT "The EphA4 and EphA1 receptor tyrosine kinases and ephrin-B2 ligand
RT regulate targeted migration of branchial neural crest cells.";
RL Chr. Biol. 7:561-570(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20099673;
RA Helbling P.M., Brandt A.W.;
RA Wilkinson D.G., Saulnier D.M., Robinson V., Christiansen J.H.;
RT "Comparative analysis of embryonic gene expression defines potential
RT interaction sites for Xenopus EphA4 receptors with ephrin-B ligands.";
RL Dev. Dyn. 216:361-373(1999).
DR EMBL; AF128844; AAD32610.1;
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR ProDom; PD002533; Ephrin; 1.
FT NON TER
SQ SEQUENCE 205 AA; 22256 MW; 7DEDD34277260C87 CRC64;

Query Match 13.5%; Score 331; DB 13; Length 205;
Best Local Similarity 36.4%; Pred. No. 2.2e-18;
Matches 82; Conservative 35; Mismatches 72; Indels 36; Gaps 7;

QY 128 EFRSHDYIITATSDGREGLESLOGGVCLTRGKVLRYGSP-----RGAVPRKPV 182
DB 1 EFRQKDYIITATSDGREGLESLOGGVCLTRGKVLRYGSP-----RGAVPRKPV 60
QY 183 EMPN--ERDGAASLEPKGENTPGDTSNATRGAGPPLPPSPKAVAGAGLALL 240
DB 61 ESGTNGKSTTSFHYNPEGSSTEGKAGHSILGSEVAL-----FAGIASGSIIFIV 113
QY 241 GVGAGAGMCMRRRAKPSERHPPGSGFGRGSLGLG-----GGGAGPREAREPGL 293
DB 114 IITLVVLKTRRRKHSPOHT-----TSLISTATPKRGSGNNNG---SEPSDI 161
QY 294 GIALRGGAADPEFCHEKSGDYGHPIYIVDGPPOSPPNIY 338
DB 162 IIFLR---TAEVCPHYEKVSGDYGHPIYIVDGPPOSPPNIY 203

RESULT 8
Q9U3M2 PRELIMINARY; PRT; 237 AA.
AC Q9U3M2;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE C43F9.8 protein.

```

```

GN C43F9.8.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mortimore B.J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z82262; CAB54195.1;
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR ProDom; PD002533; Ephrin; 1.
SQ SEQUENCE 237 AA; 26748 MW; B9B2D9FC71FE4FC CRC64;

Query Match 8.8%; Score 214.5; DB 5; Length 237;
Best Local Similarity 25.7%; Pred. No. 3.7e-09;
Matches 53; Conservative 41; Mismatches 83; Indels 29; Gaps 6;

QY 11 VRVALLLGLGLVS--GLSLPYNNNSANKRQAG--GYLYPOIGDRLLCPARP 68
DB 1 MQIATFILSLPPIGARKIPDINWISSNPIFDVSNTEIVSVHIGDRSINCPSDER 60
QY 69 GPHSSPNTEFYKLYVGAGGRCRCEAPPNLLITCRPLDARFTTKQYSPNLMGE 128
DB 61 G-----KXESYIYVSDDEYDHCFL--SKPLVACDNQTNISINIVFSPFPGGFE 114
QY 129 FRSHDYIITATSDGREGLESLOGGVCLTRGKVLRYGSP-----RGAVPRKPV 169
DB 115 FQPKNYFLIKSEVDALLIYETANQIFPGTSBOTLEGIRKKGDLCTAKQMKIKFVQ 174
QY 170 SPRGAVPRKPVSEMPERDGAHS 195
DB 175 DRRGLENPK--FAARTLKORDAHS 198

RESULT 9
Q9U474 PRELIMINARY; PRT; 279 AA.
AC Q9U474;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE VAB-2 (Hypothetical protein Y37E11AR.6).
GN VAB-2 OR Y37E11AR.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=N2;
RX MEDLINE=20084449; PubMed=10619431;
RA Chin-Sang I.D., George S.E., Ding M., Moseley S.L., Lynch A.S.,
RA Chisholm A.D.;
RT "The ephrin VAB-2/EFN-1 functions in neuronal signaling to regulate
RT epidermal morphogenesis in C. elegans.";
RL Cell 99:761-790(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=BRI-STOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).

```



Db 614 IOTL 618

# RESULT 11

Q98721 PRELIMINARY; PRT; 202 AA.  
 ID O98721  
 AC O98721  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Ephin-A6 (Fragment).  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RX NCB  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Menzel P., Valencia F., Godement P., Dodeliet V.C., Pasquale E.B.;  
 RT "Ephin-A6, a new ligand for Epha receptors in the developing visual  
 system";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF317286; AAK00944.1; --  
 DR InterPro; IPR001759; Ephin.  
 DR Pfam; PF00812; Ephin; 1.  
 DR PRINTS; PR01347; EPHRIN.  
 DR ProDom; PD002533; Ephin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 FT NON TER  
 SQ SEQUENCE 202 AA; 22624 MW; 91E2716FF91353F9 CRC64;

Query Match 7.3%; Score 178.5; DB 13; Length 202;  
 Best Local Similarity 27.5%; Pred. No. 2.1e-06;  
 Matches 56; Conservative 27; Mismatches 84; Indels 37; Gaps 7;

QY 33 YVWNSANKRFOAGGVLVYPOIGRLDLCPRAPRPPGSHSPNYFYLLVYGAGQGRG 92  
 DB 25 YVWNSGNPRF-LDDYISIVSINDLDIYCPHYSAPTPA---SSFTLFWDEGGRGC 79  
 QY 93 EAPPAVNLITCDR--PDLDRTFKFOEYSPNMGHFRSHDYIAT-SDGTREGL 148  
 DB 80 SERPGAFRMECKPRAPVPVPRFSEKIQRTFPLGFEFRGENTYIISVPTPS---- 135  
 QY 149 ESTIQGVCLTRGKYLRLVQSPRGAVPRKVSSEMERDGAHSLPECKENLPDPT 208  
 DB 136 ---AGRCCLKRVSYCCR-----ASTPEPLTEVPNSQPRGR-----GGPE 171  
 QY 209 SNATSRGAGPLPPSPMAVAGA 232  
 DB 172 GDAAGPRDAAPIPORSRTLVALA 195

# RESULT 12

Q9D7K8 PRELIMINARY; PRT; 205 AA.  
 ID Q9D7K8  
 AC Q9D7K8  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Adult male tongue cDNA, RIKEN full-length enriched library,  
 DE clone:231000475, full insert sequence.  
 GN EFNA1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX NCB  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=TONGUE;  
 RC MEDLINE=21085660; PubMed=11217851;  
 RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Kawai T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuenl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake U., Bozell D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK009144; BAB26102.1; --  
 DR MGD; MGI:103236; Etnal.  
 DR InterPro; IPR001799; Ephin.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00812; Ephin; 1.  
 DR PRINTS; PR01347; EPHRIN.  
 DR ProDom; PD002533; Ephin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 SQ SEQUENCE 205 AA; 23772 MW; E37E55767459A4EC CRC64;

Query Match 6.8%; Score 167; DB 11; Length 205;  
 Best Local Similarity 26.1%; Pred. No. 1.7e-05;  
 Matches 43; Conservative 34; Mismatches 76; Indels 12; Gaps 4;

QY 19 LIGVLVGLSGLSIEPYWNSANKRFOAGGVLVYPOIGRLDLCPRAPRPPGSHSPNY 77  
 DB 8 LIGVCSLMAADRHIVFNMNSNPKFREE-DYTVHVLNLYLILCPHYDDSV-ADAVE 65  
 QY 78 FYLLVYGAGQGRCAAPAPNLLTCDRDL--DLRTFKFOEYSPNMGHFRSHD 134  
 DB 66 RYLLVWVHEQVYACQPSQDQVRNCKRPSAKHGEKLSVFORPTPIILGKEFEGHS 125  
 QY 135 YVLIATSDGTREGLSIOGVCLTRGKYLRLVQSPRGAVPRK 179  
 DB 126 YVYISKRYHGE-----SQCLKLVTVNGKITHRPAHVNPQE 163

# RESULT 13

Q9CZS8 PRELIMINARY; PRT; 206 AA.  
 ID Q9CZS8  
 AC Q9CZS8  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE 10 days embryo cDNA, RIKEN full-length enriched library,  
 DE clone:2610529M21, full insert sequence.  
 GN EFNA4.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX NCB  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RC MEDLINE=21085660; PubMed=11217851;  
 RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuenl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,





Db 83 KPHSFDOJLWECSEKPFAPHAPEKSEKFORFTFTLGEKESYYSI-----K 133

QY 150 SL--QGVCITRGKVTLLRVGQSPRGAVPRKPVSEMPMRDGAHSLPEKKNLPGDP 207

Db 134 PLHHGGECLRLKVDV-----GPHGSKKKKKVEKEELEGMAAGVHNPNRLPAD 188

QY 208 TSNATSRGAGPLPPSPMPAVAGAG 233

Db 189 -----PIAMIFVYQSVG 201

## RESULT 16

Q9ESQ2 PRELIMINARY; PRT; 1691 AA.

AC Q9ESQ2; 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, last annotation update)

DE Type IV collagen alpha 5 chain.

GN COL4A5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10990;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20536494; PubMed=10965041; Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R., Kishimoto Y., Sado Y., Yoshiooka H., Ninomiya Y.; "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in Epithelial Basement Membranes"; J Biochem. 128:427-434(2000).

RL EMBL; AB041350; BAB13673.1; -.

DR MGD; MGI:88456; COL4A5.

DR InterPro; IPR000087; Collagen.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD000007; Collagen; 3.

DR ProDom; PD003923; ProcollagenC4; 1.

DR SMART; SM00111; C4; 2.

KW Collagen.

SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match 6.3%; Score 153.5; DB 11; Length 1691;  
Best Local Similarity 24.8%; Pred. No. 0.0022;  
Matches 86; Conservative 29; Mismatches 119; Indels 115; Gaps 18;

QY 52 PQTGDRDLDCPRARPPPHSSPNVEFYKLYLVGAQGRCEAPAPNLTTCDRPDL 111

Db 442 PQLBSR-DEIC-KAGPPGPPGPGDK-----GLQGERGVGDKDPTCENC----- 484

QY 112 RFTIKFOEYSPNLMGHEFRSHDYIIATSDGREGLSIQ-----GYCLTRGMKV 163

Db 485 -----IGTGISGPPGPGPLPGPPSGIGIPGEGDK- 517

QY 164 LLRVGQS-----PRG-GAVPRKP-----YSEMP-MERDRG-----AAHSL- 196

Db 518 -----GAGITGPPKPLPDPGAPGPGSKGPPGDLTLPKMKGKXGELGPPGAGL 573

QY 197 ---EPKKNLPGDP-----TSNATSRGAGPLPPSPMPAVAGAGLALL----- 238

Db 574 LPSPGKDGPLPLPGPKGSPGKITFKGERGPPGSPGLPLPGNMGPPTGVPFGPPPIGE 633

QY 239 --LLGVGAGAGMCKRRRAKPSSES--RHNG-----PGSPFGSGSLGCGGGMCKRRRAERG 291

Db 634 KGIGGVAGNPGQGLPEPKADPGQITTPGKPKGLPGSPGSDGSLVLPDGLPGQGLPG 693

QY 292 ELGI-----ALRGGAADPPCPHYEKVSGD---YGHFVYIVODGPPQSP 333

Db 694 ILGSKGEPGIGTGPPEPPPKGPPGPIGPGPAGAPGAPGRMGPPGPPGP 742

## RESULT 17

Q9IB91 PRELIMINARY; PRT; 1447 AA.

AC Q9IB91; 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, last annotation update)

DE Type I collagen alpha 1.

GN COL1A1.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RA Goto T., Katada T., Kinoshita T., Kubota H.Y.; "Expression and characterization of Xenopus type I collagen alpha 1 (COL1A1) during embryonic development."; Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RL EMBL; AB034701; BAA94972.1; -.

DR InterPro; IPR000085; Fib collagen\_C.

DR InterPro; IPR001007; VWF\_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR Pfam; PF00093; VWC; 1.

DR ProDom; PD000007; Collagen; 2.

DR ProDom; PD002078; Fib collagen\_C; 1.

DR SMART; SM00038; COLFI\_1.

DR SMART; SM00214; VWC; 1.

DR PROSITE; PS01206; VWF\_C; 1.

KW Collagen.

SQ SEQUENCE 1447 AA; 137446 MW; AAA6DD2B4158E38B CRC64;

Query Match 6.2%; Score 151.5; DB 13; Length 1447;  
Best Local Similarity 27.2%; Pred. No. 0.0026;  
Matches 86; Conservative 13; Mismatches 128; Indels 89; Gaps 16;

QY 67 PPGHSSPNVEFYKLYLVG-----GAQGRCE-APPANLTTCDRPDL 111

Db 756 PPGPGAGGDK-----GEXGAPGAPGTGAGGAPGERGEPAAPGAPG--AGPPGAD- 805

QY 112 RFTIKFOEYSPNLMGHEFRSHDYIIATSDGREGLSIQ-----GYCLTRGMKV 162

Db 806 -----GQGARKEQDSDG-----AKGDAPSGPLPPTGAPGAGLSPKARG 851

QY 163 -----VLLRVG-OSPRGAVPRKPVSEMPMERDRG-----AAHSLPEKEN 202

Db 852 APGPFGATGFGAAGRVGPPGSGNAGPPGSPGAKGKAPRGRTGPAGRSGEPGAG 911

QY 203 LPDPTSNATSRGAGPLPPSPMPAVAGAGLALL-LLVGAAGMCKRRRAKPSSES 261

Db 912 PPGPPGEGK-SPGSDGPGAPGIPGPGVAGSRGTVALPMMRGGRGSGLPGPAGEGKQ 970

QY 262 RHPP-----GSFGSGSLGCGGGMCKRRRAEPPELIALRGGAADPPCPHYEKVSGD 317

Db 971 GSSGSPGGRGPPGSPGPGIGGPPGSGRGAOSSEBAPGR-DGAVP-----KGD 1020

QY 318 YGHFVYIVODGPPQSP 333

Db 1021 RGEAA-----GPPGAP 1031

## RESULT 18

Q17036 PRELIMINARY; PRT; 325 AA.

AC Q17036; 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, last annotation update)

DE Hypothetical 31.6 kDa protein.

GN T15B7.4.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OX NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Pauley A., Gatlung S.;  
 RT "The sequence of C. elegans cosmid T1587";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022985; AAB69961.1.-.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col cuticle\_N.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF01484; Col cuticle\_N; 1.  
 DR Hypothetical protein.  
 KW SEQUENCE 325 AA; 31591 MW; 6538D4E37C9C202 CRC64;  
 SQ  
 Query Match 6.1%; Score 149; DB 5; Length 325;  
 Best Local Similarity 26.0%; Pred. No. 0.00074;  
 Matches 92; Conservative 25; Mismatches 109; Indels 128; Gaps 23;  
 QY 13 VGALLLGLGVSGLSL-----EPVYMSANKRFQAGE-----YVLPGIGRLDLC 62  
 DB 1 NSASTLVTVASAGSIALVVCFTVGMIFNDINSFYDEKIELKEFKYECIA--KQAMT 58  
 QY 63 PPARPPGHSFNFYFYLYLVG-----GNQSRCEA-PPAPNLATTCDRPDL 111  
 DB 59 PTTRSSSS-----FLGRMKQAQNCAGQSGCAPGPGP----- 96  
 QY 112 RTTFQGVSPVLMGHERSHHDYITLSDGTRBGLSLG--GVCITRCKVTLRVQ 169  
 DB 97 ----DGQGAPEQGHG-----LAQPGSGARINPATGRPGFCT----- 133  
 QY 170 SPRGAVPRKPVSEMPERDRGAHSLPEGKENTLPGDPTSNATSGAEGPLPPSPMPAVA 229  
 DB 134 CPAGAPGAPG-----PGA-----PGRKANNQPGGAPAGS--GGRGP--A 175  
 QY 230 GAAGLALLLGVAGAGAMCWRRRRAKPSRRHPG--PGSFGRGQ--SLGIGGGGEM-- 283  
 DB 176 GDAG-----SPGQPGHFGSGPGRGGRGGRGRLPGSGRPGP 212  
 QY 284 -GPRAEPRGELIALRGGAADP--PFCPHYKVSQDVCHPYIYVQDPPQSPN 335  
 DB 213 PGP-AGGFGQPG--HSGGASPGPGP-----PSPSQPGHSGNDVPGAPGN 257  
 RESULT 19  
 Q17208 PRELIMINARY; PRT; 569 AA.  
 AC Q17208;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Collagen (Fragment).  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Dictyoptera;  
 OC Bombycoidea; Bombycidae; Bombyx.

OX NCB1\_TaxID=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EUROPEAN 200X300; TISSUE=IMAGINAL WING DISC;  
 RX MEDLINE=97148965; PubMed=8995790;  
 RA Chareyre P.P., Beeson M.M., Fourche J.J., Bosquet G.G.;  
 RT "Identification of a Bombyx collagenous protein with multiple short  
 RT domains of Gly-Xaa-Yaa repeats: cDNA characterization and regulation  
 RT of expression";  
 RL Insect Biochem. Mol. Biol. 26:677-685 (1996).  
 DR EMBL; Z30348; CAA83002.1.-.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001419; Glutennin.  
 DR Pfam; PF01391; Collagen; 5.  
 DR PRINTS; PR00210; GLUTENIN.  
 FT NON TER 1  
 SQ SEQUENCE 569 AA; 56583 MW; FE371482F0A3B0AF CRC64;  
 Query Match 6.1%; Score 149; DB 5; Length 569;  
 Best Local Similarity 29.3%; Pred. No. 0.0014;  
 Matches 51; Conservative 14; Mismatches 77; Indels 32; Gaps 6;  
 QY 171 PRGAVPRKPVSEMPERDRGAHSLPEGKENTLPGDPTSNATSGAEGPLPPSPMPAVAG 230  
 DB 1 PGQGGPIKPGQ-----PGYRQPGQPGYPGQ--GQPAAPGQPGQPGQPGQPTPG 50  
 QY 231 AAGLALLLGVAGAGAMCWRRRRAKPSRRHPG-----PGSFGRGSLGLGGGGM 283  
 DB 51 QAG-----QPGYRQGG-----QPIKPAQPGHPGQPGQPGYRPGQPGQPGYRPGQ 99  
 QY 284 GPRAEPRGELIALRG-----GGAADPPFCPHYKVSQDVCHPYIYVQDPPQSPN 333  
 DB 100 PGQPGPGQGGYRPGQPGQPGQPGQPGQPGYRPGQSGQPGYRPGQPGQPGQ 153  
 RESULT 20  
 Q19111 PRELIMINARY; PRT; 316 AA.  
 AC Q19111;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE F02D10.1 protein.  
 GN F02D10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OX NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Swidburne U.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology";  
 RL Science 282:2012-2018 (1998).  
 DR EMBL; Z67990; CAA91932.1.-.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col cuticle\_N.  
 DR Pfam; PF01391; Collagen; 2.  
 DR Pfam; PF01484; Col cuticle\_N; 1.  
 SQ SEQUENCE 316 AA; 29444 MW; 4D76D5BA07923499 CRC64;  
 Query Match 6.1%; Score 148.5; DB 5; Length 316;  
 Best Local Similarity 28.5%; Pred. No. 0.00078;  
 Matches 49; Conservative 12; Mismatches 78; Indels 33; Gaps 6;  
 QY 168 GSPRGAVPRKPVSEMPERDRGAHSLPEGKENTLPGDPTSNATSGA 216  
 DB 144 GASGKAAPPCSPTRPPQ--PCPAPPGPPGPDGTGPGGPGGAPGASGSGRPGAP 202

```

QY 217 EGPLPSPMPAVAGAGLALLLVAGAGAACMRRAKPSSESRHGP-GSFGRGSL 275
DB 203 PGPAGPNDGQPGQPGG-ODGASSAGG-----EKGPPAPPPGAGAGPDGGS 253
QY 276 GLGGGGMGPR-----EAPGELGALRGGAADPPPCPHYEKVSG 316
DB 254 GSGAGPGRGKPPGAGQPGSDGNPETAAPGNPGEGEKICPKKCALDG 305

RESULT 21
Q61434
ID 061434 PRELIMINARY; PRT; 1140 AA.
AC 061434;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioke H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL: D17546; BA04483.1; -.
DR HSSP: P39061; IKOB.
DR MGD: MGI:88449; Col15a1.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 8.
DR Pfam: PF02210; TSPY; 1.
DR ProDom: PD000007; Collagen; 1.
FT NON_TER 1
SQ SEQUENCE 1140 AA; 11516 MW; 8B0C7E6862B3BDFE CRC64;

Query Match 6.0%; Score 146.5; DB 11; Length 1140;
Best Local Similarity 23.7%; Pred. No. 0.0049;
Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

```

```

RESULT 22
Q91718 PRELIMINARY; PRT; 1491 AA.
ID Q91718
AC Q91718;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Alpha-1 type II collagen.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92011898; PubMed=1918153;
RA Su M.W., Suzuki H.R., Bieker J.U., Solursh M., Ramirez F.;
RT "Expression of two nonallelic type II procollagen genes during Xenopus
RT laevis embryogenesis is characterized by stage-specific production of
RT alternatively spliced transcripts."
RL J. Cell Biol. 115:565-575(1991).
DR EMBL: M63596; AAA49679.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_Collagen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF00093; VWC; 1.
DR ProDom: PD000007; Collagen; 3.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SW00214; VWC; 1.
DR PROSITE: PS01208; VWF; 1.
DR Collagen.
KV SEQUENCE 1491 AA; 142495 MW; 43026FF08FE0314 CRC64;

Query Match 6.0%; Score 146.5; DB 13; Length 1491;
Best Local Similarity 31.1%; Pred. No. 0.0067;
Matches 56; Conservative 11; Mismatches 80; Indels 33; Gaps 8;

```

RT partial structure of the corresponding gene, and comparison of the  
 RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen  
 RT chain."  
 RL J. Biol. Chem. 269:13929-13935 (1994).  
 RN [2]  
 RP SEQUENCE OF 1-562 FROM N.A.  
 RX MEDLINE=94240112; PubMed=8183894;  
 RA Rehn M., Pihlajaniemi T.,  
 RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the  
 RT collagenous sequence, a distinct tissue distribution, and homology  
 RT with type XV collagen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238 (1994).  
 RN [3]  
 RP SEQUENCE OF 1-562 FROM N.A.  
 RX MEDLINE=95181468; PubMed=7876242;  
 RA Rehn M., Pihlajaniemi T.,  
 RT "Identification of three N-terminal ends of type XVIII collagen chains  
 RT and tissue-specific differences in the expression of the corresponding  
 RT transcripts. The longest form contains a novel motif homologous to rat  
 RT and Drosophila frizzled proteins."  
 RL J. Biol. Chem. 270:4705-4711 (1995).  
 DR EMBL; U03715; AAC52903.1; -  
 DR EMBL; U03716; AAC52903.1; JOINED.  
 DR EMBL; U03718; AAC52903.1; JOINED.  
 DR EMBL; U34607; AAC52903.1; JOINED.  
 DR EMBL; U34608; AAC52903.1; JOINED.  
 DR EMBL; U34609; AAC52903.1; JOINED.  
 DR EMBL; U34610; AAC52903.1; JOINED.  
 DR EMBL; U34611; AAC52903.1; JOINED.  
 DR EMBL; U34612; AAC52903.1; JOINED.  
 DR EMBL; U34613; AAC52903.1; JOINED.  
 DR EMBL; U11637; AAC52179.1; -  
 DR HSSP; P39061; 1KOE.  
 DR MGP; MGI:88451; Col18a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000024; Fz\_domain.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 8.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM0063; FRI; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.  
 DR PROSITE; PS0038; Fz; 1.  
 KW Signal.  
 SQ SEQUENCE 1774 AA; 182229 MW; CP4D9BC9388EF232 CRC64;  
 Query Match 6.0%; Score 146.5; DB 11; Length 1774;  
 Best Local Similarity 23.7%; Pred. No. 0.0081;  
 Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;  
 QY 2 GPHSGGPGVAVGALLIGVIG-LVSGLSLEPVYVWNSANKRFQEGGVLYLPQIGDLDL 60  
 DB 798 GPP--GPGPQ-----GPGAPVVGQSPNSQPV----- 821  
 QY 61 LCPRAR-PPGPHSSPNVEFYLYLVGAQGRRC----- 93  
 DB 822 --PGAQGPFGPGPGPGKO-----GTPGRDGEFGDPRDPRDGPQGPPTPDVG 871  
 QY 94 -----APPANLLITCDRPLDLRFTIKFOEYSPNMGHEFRSHHYIAT 140  
 DB 872 PKGERGDPGIGRGGPGP-----PGPPGSPFRQDKLTFIDME 908  
 QY 141 SDGTGEGHESIQGVCLTRGMKVLIRVQSGPRGAVPRKP--VSEMPMERDR-GAASLE 197  
 DB 909 GSGFGSDLESARG-----PRGFRGPPGPGVGLGCEPERFINSYA 951  
 QY 198 PKEKENVLPGDPTSNATSRGAEGP--LP-PPSMPAVAGAAAGLALLILGYAGAGAMCMRRR 254  
 DB 952 PGPAGLPGVP-----GKEGPPGPGPPGPPGPGPGKGP-----PGVAGKQGSV----- 994

QY 255 RAPSSESRHPP-GSPFGSGSLGCGGGMREAPFELGIALRG--GAADPEFCPY 311  
 DB 995 ---GDVGI.PGPKSGKGDIGPFGKSGLA---GSPFVGPFGPPGPPGPPGPAAGF 1047  
 QY 312 EKYSGDYGHVYIVQ-----DGPPOSF 333  
 DB 1048 DDMESG-GIPLWTTARASDGLQGPFGSP 1074  
 RESULT 24  
 Q90800  
 ID 090800 PRELIMINARY; PRT; 675 AA.  
 AC Q90800  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Collagen-alpha-3 type IX precursor.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CARTILAGE;  
 RX MEDLINE=92241276; PubMed=1572350;  
 RA Brewton R.G., Oupenekata M.V., van der Rest M., Mayne R.;  
 RT "Cloning of the chicken alpha 3(IX) collagen chain completes the  
 RT primary structure of type IX collagen."  
 RL Eur. J. Biochem. 205:443-449 (1992).  
 DR EMBL; X64712; CAA45967.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 11.  
 DR ProDom; PD000007; Collagen; 3.  
 KW Collagen; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 675 COLLAGEN-ALPHA-3 TYPE IX.  
 SQ SEQUENCE 675 AA; 63069 MW; 95094637A339D72D CRC64;  
 Query Match 6.0%; Score 146; DB 13; Length 675;  
 Best Local Similarity 28.8%; Pred. No. 0.0029;  
 Matches 64; Conservative 9; Mismatches 71; Indels 78; Gaps 12;  
 QY 166 RVGQSPRGAVPRKPYSEMPMERDRGAASLEPGKENTPGDTSNATSRGAPLPSPSM 225  
 DB 23 RVG--PQGPFGPRGPPGPGKGDIDG-----EPGSGLPGPP-----GPGAPGX 65  
 QY 226 PAVAGAG--GLALLILGYAGAGAMCMRRRAKSESRRHP-----PGSFGRG- 272  
 DB 66 PGAAGBAAGLPGLP---GVDGLTGT-----DPPGPPGPPDRGALGPAGPPGAGKL 115  
 QY 273 -----GSLIGLGGGGM-----GPR-----EAPGELGIALRG----- 300  
 DB 116 PGPFGPPGSGLP.GGAGFRGPPGPGSGLP.GFP.PP.PP.PPGLAGIIPREGGLQCPALCP 175  
 QY 301 -GAADPPFCPHYKXVSGDYGHVYIVQGP-----PQSPENI 336  
 DB 176 PGPPGPPGPMGPKGHTGKGEPEGIGKEGKSPGPPPGPI 217  
 RESULT 25  
 Q90250  
 ID Q90250 PRELIMINARY; PRT; 1669 AA.  
 AC Q90250  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Alpha 3 collagen IV.  
 GN COL4A3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RX MEDLINE=2000534; PubMed=10534397;  
 RA Lu W., Phillips C.L., Kilen P.D., Hlaing T., Harrison W.R.,  
 Elder F.B., Miner J.H., Overbeek P.A., Weiler W.H.,  
 RT "Insertional mutation of the collagen genes *col4a* and *col4a4* in a  
 RT mouse model of alport syndrome.";  
 RL Genomics 61:113-124(1999).  
 DR EMBL: AF169387; AAD50449.1;  
 DR MGD; MGI:104686; Col4a3.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001442; ProcollagenC4.  
 DR InterPro: IPR000504; RNA\_rec\_mct.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 21.  
 DR Prodom: PD000007; Collagen; 6.  
 DR Prodom: PD003923; ProcollagenC4; 1.  
 DR SMART: SM00111; C4; 2.  
 DR SMART: SM00030; RNP\_1; UNKNOWN\_1.  
 DR PROSITE: PS00030; RNP\_1; UNKNOWN\_1.  
 KW Collagen.  
 SQ SEQUENCE 1669 AA; 161769 MW; 3097659739447B2 CRC64;

Query Match 6.0%; Score 146; DB 11; Length 1669;  
 Best Local Similarity 24.3%; Pred. No. 0.0083;  
 Matches 97; Conservative 34; Mismatches 134; Indels 134; Gaps 22;

QY 4 PHSGPGVAVGALLLGLVGLV-----VSGLSLEP-----VYNSANKRFOAEGY 48  
 DB 1008 FHGMPS-----MGIMVPGPKGRKGTSLGLPGLAGPRGTHGPDK---GEPGV 1056  
 QY 49 VLVQIGRLDLCPRRAPRPGHSSPNYERYKLYLVGAGGRCE----- 93  
 DB 1057 SEGAPG-----PPGPKGDPG-----LPDGKKGKGERGVPPGPGQSGPAG 1097  
 QY 94 --APPAPVLLTCDRP---DLDRFTIKFOEYSPNLMGHEFRSHHDYIATSDGTRE 147  
 DB 1098 DGAGSGESGSPGPKPPGADLGL---KQKRGFPFG-----STGPGPG 1141  
 QY 148 LESLQGGVCL-----TRGMKYLRAVGSGPRGAGVPRKVSSEMERDRGA 193  
 DB 1142 LPGLPFGPMKRGDGRDGIPPRPGKEGTGLGAYPGPKGS--PGVPGA---KDRGV- 1194  
 QY 194 HSLRPGKENTPGDPTSNATSGAGCPRLPPSPMAVAGAGLALLLVGAGAGAMCMR 253  
 DB 1195 ---PGSGLPG-----KRGVMDVPGQPGPTAGLPDPGPGAIIPGPGDRGLPG 1243  
 QY 254 RRAKPSRHPG-PGSFGRG-----GSLGLGGG-GMGPREAEPGELIALRG 300  
 DB 1244 LRGNPGEGRPPGPPGKIGKIKDKGMGPPGPKGLPGTVGDMP-PGFPAGPTGLPG 1302  
 QY 301 GAAPPFGPHYKVSQDYGHPY---IVQDP--PQSP 334  
 DB 1303 VRGDPGF-PGFPGKIGKGNKGNFLGPIGHPPGVPGKPP 1340

RESULT 26  
 ID 017805 PRELIMINARY; PRT; 305 AA.  
 AC 017805;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE F14F7.1 protein.  
 GN F14F7.1  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurray A.A.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98069613; PubMed=9851916;  
 RA none,  
 RT "Genome sequence of the nematode *C. elegans*: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z81503; CAB04111.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002486; Col\_cuticle\_N.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 SQ SEQUENCE 305 AA; 29497 MW; 161BP90D024C2B6 CRC64;

Query Match 5.9%; Score 145.5; DB 5; Length 305;  
 Best Local Similarity 22.5%; Pred. No. 0.0013;  
 Matches 83; Conservative 21; Mismatches 122; Indels 143; Gaps 12;

QY 11 VRVGLLLG-----VLGLVGLSLEPYRN-----SANK----- 40  
 DB 5 VRLKAYRLAASAVAFCLVSVSCITLPMVYVSGNRVRLVQEVSEFKASAEVTEV 64  
 QY 41 -----RFOAEGGYLVYPQIGRLDLCPRARPPGHSPPNYEFTLYLVG 86  
 DB 65 NLRASATNSTRASAHAGYGY-----AQPGGGGG-----GG 97  
 QY 87 AQGRCEAPPAVNLITDRPDLRFTIKFOEYSPNLMGHEFRSHHDYIATSDGTRE 146  
 DB 98 GQCSGCCRPFGPGAGCTPKPG-----RPAFGAMG----- 128  
 QY 147 GLESLQGGVCLTRGMKYLRAVGSGPRGAVPRKVSSEMERDRGAHSLRPGKENTPGD 206  
 DB 129 -----MGNPQKGSQPCHVYVPPCKPCGG---RPGSPGPR 165  
 QY 207 PTSN-----ATSRGA-----EGPLPPSPMAVAGAGLALLLVGAGAGAMCM 251  
 DB 166 PQSDQQRPAATGGAARPPGPKGPRGAPGNSGAGAPGQNDAGYGGVGAAPG 225  
 QY 252 RRRAKPSRHPGSGSGRSGSLGLGGGGM-----GPREAEPGELIALRGGAAPPF 307  
 DB 226 AGPRAPAPGPGSGSGGPRGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 284  
 QY 308 CPHYEKVSG 316  
 DB 285 CPKYCAIDG 293

RESULT 27  
 ID 094620 PRELIMINARY; PRT; 308 AA.  
 AC 094620;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Cuticle prepocollagen.  
 GN COL-2.  
 OS Meloidogynae incognita (southern root-knot nematode).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
 OC Tylenchidae; Heteroderidae; Meloidogyninae; Meloidogynae.  
 OX NCBI\_TaxID=6306;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CUTICLE;  
 RX MEDLINE=98324411; PubMed=9662034;  
 RA Wang T., Deem C.M., Huesey R.S.;  
 RT "Identification of a Meloidogynae incognita cuticle collagen gene and  
 RT characterization of the developmental expression of three collagen  
 RT genes in parasitic stages.";  
 RL Mol. Biochem. Parasitol. 93:131-134(1998).  
 DR EMBL; U68729; AAC48358.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002486; Col\_cuticle\_N.  
 DR Pfam; PF01391; Collagen; 2.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.

```

KW Collagen
FT CHAIN 80 308 COLLAGEN
SQ SEQUENCE 308 AA, 29361 MW, 825AEF0249FDC78 CRC64,

Query Match 5.9%; Score 145.5; DB 5; Length 308;
Beat Local Similarity 28.4%; Pred. No. 0.0013;
Matches 46; Conservative 14; Mismatches 67; Indels 35; Gaps 6;

OY 171 PRGGAVPRKPVSEMPERBERGAHSLPEKCNLPQDPTSNATNRGAEGPLPPSPMAVAG 230
Db 155 PPGGQPPGQPPGPPGPGGQGGGPGQPPGQGGQSGPPEP-----GAGPGGSGAGG 204
OY 221 AAGGATLLILGVAAGAGAMCMRRRRARKEPSRRPG---PGSFGRGSGTGL-GGGGNGP 285
Db 205 QPGG-----PGQPGADA-QSQEAGPAPPPGPDGAGAGQPGNGGPGAGQGGPGGPGGP 255
OY 266 R-----EAERPGELIALRGGAADPFPCPYKAVSG 316
Db 256 KPPPGPPGQPPGNDGSGPGQPPGPPGNGSGGSEKGICTKCAIDG 297

RESULT 28
O9DD02 PRELIMINARY; PRT: 680 AA.
O9DD02
ID 09DD02 01-JUN-2001 (TREMBlrel. 17, Created)
AC 09DD02 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Procollagen, type IX, alpha 1.
GN COL9A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851,
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Caavaant T.,
RA Laetschmann W., Gaesteland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikiado I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Stuehl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarini R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima Y., Mazzarelli U., Mombaerts P.,
RA Nordone P., Ring B., Schenbach M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Willecker C., Wilming L.,
RA Wyshak-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
DR EMBL; AK011547; BAB27690.1; -.
DR MGD; MGI:88465; Col9a1.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 11.
DR PRINTS; PR01582; KV33CHANNEL.
DR Prodom; PD000007; Collagen; 3.
KW Collagen.
SQ SEQUENCE 680 AA, 64733 MW, 5C0FB3B9892B4D91 CRC64;

Query Match 5.9%; Score 145; DB 11; Length 680;
Beat Local Similarity 24.0%; Pred. No. 0.0035;
Matches 87; Conservative 22; Mismatches 131; Indels 122; Gaps 17;

```

```

Db      150 GPP-GPFGPS-----GTIGFHNG-----DPL 169
QY      62 CPRAAPPGPHSSPNVEFYKLVIVG-----GAQGRCEAPAPADNLLITCDRDLDIRFTIK 116
Db      170 CPNSCPSPASGYPG-----LPGRGCHKGAKGEI----- 199
QY      117 FQVSPFNLMWGFHFRSHNDHYIIATSDGREGLSEIQ--GVCLTRGMKVILRVGS PRGG 174
Db      200 -----FGRQGHKGESGDDELGEVDQGPRPGQGLRGITGITVGDXGKEKARFPDDEPGQG 254
QY      175 AVP-----RKPVSEMPMERDRGAHSLBEGKENLPDDPTSNATS--RGAEGPLPPS 224
Db      255 GLPGAAGDGQGRGPPEGTEGPEEDRGDI-----QSRRGIPGSPPKKDGTGLPGVDGDDGIPIG 309
QY      225 MPVAVGANG-----GLAALL--LLGVAAGAGAMCWMRRRAKPSE-----SRHPG--- 265
Db      310 MPTGTGBAGKRPDPDVVLQGLPGVPGLPGAKVAVAGEKNTGAPFKPOLGSSGKRGQQG 369
QY      266 -PGSGFGQSGISGL-QGGGAMGPFREAE-PBGIGIALRGQ--GAADPPCPHYEKVSGDYG 319
Db      370 PPGEVGPGRGPRLPGSRPVPGBEGSPGCIIPGKLGSVGSFGLPGLPGLPGMKGDRCGVFG 429
QY      320 HP 321
Db      430 EP 431

RESULT 29
Q9NZO6          PRELIMINARY;           FR1, 1745 AA.
AC              Q9NZO6;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Collagen type V alpha 3 chain.
GN      COL5A3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mamalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=PLACENTA, AND HEART;
RX      MEDLINE=20187594, PubMed=10722718;
RA      Iimamura Y., Scott J.C., Greenspan D.S.;
RT      "The pro-alpha3(V) collagen chain. Complete primary structure,
RT      expression domains in adult and developing tissues, and comparison to
RT      the structures and expression domains of the other types V and XI
RT      procollagen chains.";
RL      J. Biol. Chem. 275:8748-8759(2000).
DR      EMBL; AF117941; AAF59902.1; -.
DR      InterPro; IPRO00087; Collagen.
DR      InterPro; IPRO00885; Fib.collagen_C.
DR      InterPro; IPRO01791; Laminin_G.
DR      InterPro; IPRO03129; TSPN.
DR      Pfam; PF01410; COLFI; 1..
DR      Pfam; PF01391; Collagen; 17.
DR      Pfam; PF02210; TSPN; 1.
DR      ProDom; PD000007; Collagen; 3.
DR      ProDom; PD002078; Fib.collagen_C; 1.
DR      SMART; SMO0038; COLFI; 1.
DR      SMART; SMO0282; Lamc; 1.
DR      SMART; SMO0210; TSPN; 1.
SQ      SEQUENCE 1745 AA; 172051 MW; 5E8FF97135397AAC1 CRC64;

Query Match      5.9%; Score 145; DB 4; Length 1745;
Best Local Similarity 24.0%; Pred. No. 0.01;
Matches      88; Conservative 24; Mismatches 103; Indels 152; Gaps 20

QY      2 GPP-HSGGGRVAGVALLLGVLGLVSGLSLEBPVYNANKRFQLEGYVLYEQIDGRIDL 60
Db      884 GPFGHGKMD--RPGHFGRGELG-----FQGGTG----- 911

```



```

QY 61 LCPRRPGRHSSPNVFFYKLYLVGAGQRCE-----APPAPNLLTCORPDILAE 113
Db 912 -----PPGAG-----VLGPGKTKGEVPLGERGPPG----- 939
QY 114 TIKFOYSPLMWHGHEFRSHHDVYIATSDTRGLESILQGVCVLTGMKYLKLVGQSPRG 173
Db 940 -----PPGPG-----EGGLPGLGE-----REGARG 959
QY 174 GAVPRKVSSEMPERDRGAHSLERPKENLPDPTSNATSRGAELPLP-----PSMFA 227
Db 960 ELGPPGPIGKEGPAQLRGF-----PGPKGPGDPGPTGL-KGDKGPPGVANGSPGERG 1013
QY 228 VAGAAGG-ALLLLGVAGAGAMCMRRRAKPSERHPPG-SPFRGSGSLG-LGGGGGMP 285
Db 1014 PLGPAAGIG--LPGQSSSEPPVPAKKSRRGRRPPG-TPKDGITGPIGLGPPGAAGP 1071
QY 286 -REAPGELGIALRG-----GAADPPFCPHYEKVSGDYGHVYIVODGP----- 330
Db 1072 SGEGGDGVDGAPGKHGSKDKDAGPPGQ--GIRGPAHFP-----GPGADGAGQR 1122
QY 331 QSPENY 337
Db 1123 RGPGLF 1129

```

```

RESULT 30
ID 025466 PRELIMINARY, PRT: 309 AA.
AC 025466;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COL-1.
GN COL-1.
OS Meloidogyne incognita (southern root-knot nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Meloidogyinae; Meloidogyne.
OX NCBI_TaxID=6306;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9716124; PubMed=9010847;
RA Ray C., Wang T.Y., Huesey R.S.;
RT "Identification and characterization of the Meloidogyne incognita coil
RT cuticle collagen gene.";
RL Mol. Biochem. Parasitol. 83:121-124 (1996).
DR EMBL; U40766; AAC47437.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen_2.
DR Pfam; PF01484; Col_cuticle_N_1.
SQ SEQUENCE 309 AA; 29472 MW; 0C6C9D5BA76C8916 CRC64;

```

```

Query Match 5.9%; Score 144.5; DB 5; Length 309;
Best local similarity 28.4%; Pred. No. 0.0016;
Matches 46; Conservative 14; Mismatches 67; Indels 35; Gaps 6;

```

```

QY 171 PRGAVNRKVSSEMPERDRGAHSLERPKENLPDPTSNATSRGAELPSPSMAYVAG 230
Db 156 PCPFGGQGPQPPGPPGQPGGPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPG 205
QY 231 AAGGALLLLGVAGAGAMCMRRRAKPSERHPPG--PGSFRGSGSLG-LGGGGGMP 285
Db 206 QPGG-----PGQPSPA-QSQEAGPAPPPGPDAGAGGQGNRQPAQPGGPGGPPG 256
QY 286 R-----EAPGELGIALRGGAADPPFCPHYEKVSG 316
Db 257 KGPPPGPGPDGAPGQPGTPTGPGAGAKGICPKYCAIDG 298

```

```

RESULT 31
ID 096083 PRELIMINARY, PRT: 1347 AA.
AC 096083;

```

```

DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Type V procollagen alpha 2 chain (Fragment).
GN COL5A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21451029; PubMed=11566270;
RA Valjalkila M., Melkonian M., Kvist L., Kiviemi H., Tromp G.,
RA Ala-Kotko L.;
RT "Genomic organization of the human COL3A1 and COL5A2 genes: COL5A2 has
RT evolved differently than the other minor fibrillar collagen genes.";
RL Matrix Biol. 20:357-366 (2001).
DR EMBL; AY016295; AL13166.1; JOINED.
DR EMBL; AY016290; AL13166.1; JOINED.
DR EMBL; AY016293; AL13166.1; JOINED.
DR EMBL; AY016282; AL13166.1; JOINED.
DR EMBL; AY016293; AL13166.1; JOINED.
DR EMBL; AY016294; AL13166.1; JOINED.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLF; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Collagen; 6.
DR ProDom; PD002078; Fib_collagen_C; 1.
KW Collagen.
FT NON TER 1 1
SQ SEQUENCE 1347 AA; 128822 MW; F25F20E6B23A17C CRC64;

```

```

Query Match 5.9%; Score 144.5; DB 4; Length 1347;
Best local similarity 33.0%; Pred. No. 0.0085; 77; Indels 35; Gaps 11;
Matches 61; Conservative 12; Mismatches 77; Indels 35; Gaps 11;

```

```

QY 169 QSPRG--GAV-PRKVSSEMPERDRGAHSLERPKENLPG-----DPTSNATSRGAE 217
Db 351 RGPGRDPGTVPAPGVBERGAPGNRGF-----PQSDLPKPKAGQGRGPGVSGPGSQ 405
QY 218 GPLPPSPNPAVAGAGLALLLVAGAGAMCMRRRAKPSERHPPG-SPFRG-RG--G 273
Db 406 GPPRPGPPGPGARG-----LTGNPVQGPGEKLGPLGAPGEGRGPPGPGSITIRQDPG 460
QY 274 SLGAGGGGMPREAPGELGIA-----LRGGAADPPFCPHYEKVSGDYGHVYIVODGP 329
Db 461 SMGLPGRPGSSGDDPGGEAGNAGVPGKDGAVGP-----SGPVPPGLAGRG- 514
QY 330 PQSP 334
Db 515 EQGPP 519

```

```

RESULT 32
ID 090W37 PRELIMINARY, PRT: 1420 AA.
AC 090W37;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Alpha 1 type IIR collagen precursor.
GN COL2A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=STERNAL;
RA Caixia X., Yongzhi X., Siqi G., Yiyang S.;

```

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY046949; AAK98621.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR ProDom; PD000007; Collagen; 6.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.  
 DR Collagen; Signal.  
 FT SIGNAL  
 SQ SEQUENCE 1420 AA; 134999 MW; 88D9AAB17F214FF5 CRC64;

Query Match 5.9%; Score 144.5; DB 13; Length 1420;  
 Best Local Similarity 28.1%; Pred. No. 0.009;  
 Matches 65; Conservative 11; Mismatches 90; Indels 65; Gaps 10;

QY 147 GLESLQG--GVCITRGMYLRLVQSPRGAVPRKP-----VSEMPMR----- 188  
 DB 644 GAQGLQGRRLPTGT-----DGRKATGPAQPNAGQPPGLQGRPSRGAAGIAGL 696  
 QY 189 --DRGAASLEPKENLTGDDPTSNATSRGAPLPPEPMPAVAGAGLALLIGVAGAG 246  
 DB 697 KQDRDVG--EKGPFGAGG---KDGAGLTGPIGPPEPAPGPNGRKESG--PPGSGAA 748  
 QY 247 GACMRBRRAKPSERHNG-PGSPRGSLGIGGGGGMGPRAEPBELGIALRG----- 300  
 DB 749 GA-----RGAFGERGERPAQPAQPAQPPGADGQPGAKGQGBQKDAKAPGQGPS 802  
 QY 301 -----GAADPPCPHYEKVSGDYGHVYIVDGPDPSP 333  
 DB 803 GAFGPGPTGVTGPKARGAQGPFATGFPGAAGRVGPPGNGNPPPGP 853

## RESULT 33

QY 09F342 PRELIMINARY; PRT; 775 AA.  
 AC 09F342;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein SC05128.  
 GN SC05128 OR SC9B12.13.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.  
 CX NCBI\_TaxID=1902;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,  
 RA Huang C.-H., Kleser I., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Sanders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL391751; CAC05758.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 775 AA; 75820 MW; E5032698A20A0711 CRC64;

Query Match 5.9%; Score 143.5; DB 16; Length 775;  
 Best Local Similarity 23.5%; Pred. No. 0.0054;  
 Matches 94; Conservative 23; Mismatches 130; Indels 153; Gaps 19;

QY 3 PPHSGPG-----GVRVGAALLLGVGLVSGLSLEPYWNSANKRFQAEAGGYVL 50  
 DB 133 PRKSGPGKGTGGSTNGAGLPAGSAPTAGTPGAAGAGAGAGAGAGAG----- 181

QY 51 YPQIGRLDLLCPRAPPPHSSPNYEFYKLYLVGAQ-----GRCEA---PAPNLL 101  
 DB 182 -----ARPGGSGRPGGVVSGNNAFCGARGSAGGARSATGVPVAP---- 224  
 QY 102 LTCDRDLRLFTIKQEXSPMLMGHEFHSNHDYIIANSQTRGRLBLOQGVCLTGM 161  
 DB 225 -----GH-----GGGTSFDTTEALAAPIGNGN 249  
 QY 162 KYLLRVQSPR-----GAVPRKPVSEMPERDGAASHLEPKENLPQDPTSNATSRGAE 217  
 DB 250 GNGNGSGRRPPTDAGSPRR--DDLPHYAGEGRADQAPGQQN--GNGQSPVGTGPQ 305  
 QY 218 GELPPSPMAV-----AGAAGLALLIGVAGAGAMCRRRRRAPSSSRH----- 264  
 DB 306 GPAGPPTSGPATGDSRLTPPADLGG-----LGTAGGPA-----FGTQREPAQNG 351  
 QY 265 --GPQSF--GRGSLGIGGGGGMGPRAEPBELGIALRGGAAD-----PPFCPH 310  
 DB 352 LRGPGLGPGGPGGPGGRAGSGPDAARGP-----GGGLSDDTALLTPQRPVPPGAAN 405  
 QY 311 YEKVSGD---YGHVYIVQ-----DGP-PGSPNII 336  
 DB 406 PDNISGNTVTSIPVVPGERTPAPFAGSGDGPRTTPPL 445

## RESULT 34

QY 026640 PRELIMINARY; PRT; 1747 AA.  
 AC 026640;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Alpha2(IV)-like collagen.  
 GN COLPALPHA.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidae; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
 CC Strongylocentrotus.  
 CX NCBI\_TaxID=76658;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94230414; PubMed=8175744;  
 RA Exposito J.Y., Suzuki H., Geourjon C., Garrone R., Solursh M.,  
 RA Ramirez F.;  
 RT "Identification of a cell lineage-specific gene coding for a sea  
 RT urchin alpha2(IV)-like collagen chain";  
 RL J. Biol. Chem. 269:13167-13171(1994).  
 DR EMBL; X76730; CAA54146.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 21.  
 DR ProDom; PD000007; Collagen; 6.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1747 AA; 173312 MW; EE722E878394B9B6 CRC64;

Query Match 5.8%; Score 143; DB 5; Length 1747;  
 Best Local Similarity 24.1%; Pred. No. 0.015;  
 Matches 84; Conservative 24; Mismatches 159; Indels 82; Gaps 14;

QY 7 GPGVGVGALLLGVGLVSGLSLEPYWNSANKRFQAEAGGYLVYQIGRLDLLCPRAR 66  
 DB 266 GPRGMDG---MKGATGEVGLDGSYGDEKLPVSGERGPPNGIPGLIGMNGEKOR 321  
 QY 67 PPGHSSPNYEFYKLYLVGAQGRCAAPAPNLLLTCDRDLRLFTIKQEVSPNLMG 126  
 DB 322 -DQPGRFQYDYGK--GSGDYGMNDGPEPAPDI----- 352  
 QY 127 HFRSHHDYIIATSDGREGLESLOGVCLTRGAKVLLRVQ-----SPRGAVPRKVS 182  
 DB 353 -----EIVIVGVPPQGDGPPGNGPGRGFGAIGLRDQDQPGVPGPMGGPGRGPTG 405

```

QY 183 EMWERDGAHSLPEPKENLPGDP-----TSNATSRGAGPL-----PPSPYAVAGAA 232
DB 406 SGEGRDGEGRGPRDIPFOGETGERGNDGPFGERGKXNGYBPRPBGDGRGQ 465
QY 233 GGLALLLVAGAGAGACWRRRAKPSSESNHPG-----PGSFGGSLGLGGGGGMP 285
DB 466 G-----FMGQKGRGP-----PGRAGPAGPARSGNOSFGCPGPDGTGTGLKMRGIGALG 516
QY 286 REAEPGELGALRGGAADPPPCPHYK-VSGDYGHVPIYVQGGPPSP 333
DB 517 RDGRPGSKG---ELGGIC--PCCPGIKGYPDGRGYP-----GDGSP 354

RESULT 35
076368 PRELIMINARY; PRT; 614 AA.
AC 076368;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 61.3 kDa protein.
GN F29C4.8
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2016(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Bradshaw H., Clarke K.;
RT "The sequence of C. elegans cosmid F29C4.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submision.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067616; AAC19194.1; -.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; Collagen; 2.
DR Collagen; Hypothetical protein.
KW SEQUENCE 614 AA; 61298 MW; A76E637B29CC58DA CRC64;
SQ

Query Match 5.8%; Score 142; DB 5; Length 614;
Best Local Similarity 27.2%; Pred. No. 0.0054;
Matches 87; Conservative 19; Mismatches 110; Indels 104; Gaps 21;

```

```

QY 284 -----GPREAPBELGIALRG--GAADP-----FCHYKVSQDY-----GH 320
DB 207 DGEKIRPK-----GRTSPGHDGIPGARPGGERGKGDIAFLSTYPRVASSSTASSPP 262
QY 321 FVYIVODGPSPSPNYYTS 340
DB 263 P-----GPPGVOCHAS 275

RESULT 36
096D07 PRELIMINARY; PRT; 744 AA.
ID 096D07;
AC 096D07;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 73.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Straubeberg R.;
RL EMBL; BC013581; AAH13581.1; -.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 7.
DR ProDom; PD000007; Collagen; 2.
DR PROSITE; PS01113; Clq; UNKNOWN 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 744 AA; 73364 MW; 2BC1B0955DE2C9A3 CRC64;

Query Match 5.8%; Score 142; DB 4; Length 744;
Best Local Similarity 23.3%; Pred. No. 0.0068;
Matches 72; Conservative 16; Mismatches 99; Indels 122; Gaps 13;

```

```

DE Alpha 1 (V) collagen.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031,
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BREED WHITE LEGHORN;
RX MEDLINE=20068042; PubMed=10601735;
RA Gordon M.K., Marchant J.K., Foley J.W., Igce F., Gibney E.P.,
RA Nab H.D., Barenbaum M., Myers J.C., Rodriguez B., Dabiel B.,
RA van der Rest M., Linsemeyer T.F., Upmoff W.B., Birk D.E.;
RT "Complete primary structure of the chicken alpha1(V) collagen chain.";
RL Matrix Biol. 18:481-486(1999).
DR EMBL: AF137273; AAF28099.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib collagen_C.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR001230; Premyl_site.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01410; COLFI. 1.
DR Pfam: PF01391; Collagen; 19.
DR Pfam: PF02210; TSPN; 1.
DR ProDom: PD000007; Collagen; 2.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SK00038; COLFI. 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
KM COLLAGEN.
SQ SEQUENCE 1835 AA; 184234 MW; D05BD71022D4482 CRC64;

Query Match 5.8%; Score 142; DB 13; Length 1835;
Best Local Similarity 25.7%; Pred. No. 0.019;
Matches 78; Conservative 15; Mismatches 82; Indels 128; Gaps 18;

QY 68 PGRHSSNVEYKLYLVGAGG-----RCENAPRNLLTCDDRDLDRTIFR 117
DB 566 PGRHSPGPG-----SGGLKGEVEMGPGQPRGIQGPFG-----AGK- 602
QY 118 QEVSPNLGHEFRSHDYYIATSDGTR-----EGLESIQGVCLTRGMKY 163
DB 603 -----GRRGR-----AGSDGARMPGQTGPKDNGFDGLAGLP- 636
QY 164 LIRVGQSPRGAVPRKVSSEPMERDRGAHSLPEPKENIPGPTSNATSRKAEGLPP 223
DB 637 -----EKNRGEPRGHPGPPGEGDEGEG--DGEVYGRGLPGE- 686
QY 224 SMPAVAGAGLALLLGVAAGAGCMWRRAKPSSESRHPG-PGSFGRGSLGL-GGG 281
DB 687 GPRPPEVAG-----MDGQTGPKGNV-----GPGSPGPGQGNPAGAGLPG 732
QY 282 GMGREAREPSELGIALRG-----GAADPPCFHYEKVSGDYHPYIVQDGP- 331
DB 733 PIGP-----PEKGPILGRGLPGMPGADGP-----GHP---GKEGPGKSG 773
QY 332 SPP 334
DB 774 GPP 776

```

```

OC Trichostrongyloidea; Haemonchidae; Oestertagiinae; Teladorsagia.
OX NCBI_TaxID=45464;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97039674; PubMed=8885226;
RX Johnstone I.L., Shafi Y., Mjaeed A., Barry J.D.;
RT "Cuticular collagen genes from the parasitic nematode Oestertagis
RT circumcincta.";
RL Mol. Biochem. Parasitol. 80:103-112(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Johnstone I.L.;
RN Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=20017523; PubMed=10551361;
RA Britton C., Redmond D.L., Knox D.P., McKerron J.H., Barry J.D.;
RT "Identification of promoter elements of parasite nematode genes in
RT transgenic Caenorhabditis elegans.";
RL Mol. Biochem. Parasitol. 103:171-181(1999).
DR EMBL: X96731; CA65506.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR002486; Col cuticle_N.
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF01484; Col_cuticle_N; 1.
KM COLLAGEN.
SQ SEQUENCE 284 AA; 27534 MW; 60775B7D2CC40C77 CRC64;

Query Match 5.8%; Score 141; DB 5; Length 284;
Best Local Similarity 27.2%; Pred. No. 0.0027;
Matches 61; Conservative 15; Mismatches 86; Indels 62; Gaps 12;

QY 121 SNTMGHEFRSHDYYIATSDGTRGLESIQGVCLTRGMKYLIRVGQSPRGAVPRKP 180
DB 62 SGNIRREYTRQ-----VLAAP--TRKARQ--GGGCC-----GGVSPAGPFG- 102
QY 181 VSEPMERDRGAHSLPEPKENIPGPTSNATSRKAEGLP-----PSMPAVAGAA 232
DB 103 -----PGDQGGQSDGRPGQPRNGPDGPATPAVDVPCFNCPPGPGPAGAP 151
QY 233 GLTALLLGVAAGAGCMWRRAKPSSESRHP-GPGSFRGSGSLGLGGGGMGPREAEPG 291
DB 152 GG-----RGPGSPGS-----DQPGNAGNPGGPRIPGPPGPGNAGQP- 197
QY 292 ELGILRGGAADPPCFHYEKVSGDYHPYIVQDGPSPFN 335
DB 198 APGTLTGPGSGQGPFGP-----GPGGP---GPDQPGGPGN 232

RESULT 39
Q25581 PRELIMINARY; PRT; 284 AA.
AC Q25581;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cuticular collagen.
GN COLOST-2.
OS Teladorsagia circumcincta.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Oestertagiinae; Teladorsagia.
OX NCBI_TaxID=45464;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97039674; PubMed=8885226;
RA Johnstone I.L., Shafi Y., Mjaeed A., Barry J.D.;
RT "Cuticular collagen genes from the parasitic nematode Oestertagis
RT circumcincta.";
RL Mol. Biochem. Parasitol. 80:103-112(1996).
DR EMBL: X96732; CA65507.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR002486; Col_cuticle_N.
DR Pfam: PF01391; Collagen; 3.

```

DR Pfam: PF01484; Col\_cuticle\_N; 1.  
 KW Collagen. 284 AA; 27560 MW; 7876BBD2DD41D76 CRC64;  
 SQ SEQUENCE

Query Match  
 Best Local Similarity 27.2%; Pred. No. 0.0027;  
 Matches 61; Conservative 15; Mismatches 86; Indels 62; Gaps 12;

QY 121 SPULMGHEFHSHHDYIIATSDGTREGLESLOGVCLTRGMKYLRLVSGSPRGAVPRKPE 160  
 DB 62 SGIMWEVTRTQ-----VLAAP--TRKARQS-GGGCC-----GCGVSPAGPPGP 102

QY 181 VSEMPMERDRGAASLEPKENLPCDPTSNATSRGAEGPLP-----PPSMPAVAGA 232  
 DB 103 -----PGDQGPSSDGRPGQGRNNGPPGPATAPDVRPCFNCPGPAPAGAP 151

QY 233 GGLALLLLGVAGAGACWRRRAKPSBSRHPG-PGSGFRGSLGLGGGCGMPREAEFG 291  
 DB 152 GG-----RGPQSGSPGS-----DGQPGNAGNPGGPGPIGPBPAGAGQPG--NPGSPG 197

QY 292 ELGIALRGGAADPPCFPHYEKVSGDYGHPIVVDGPPQSPPN 335  
 DB 198 APTLITGPGSGPSPGP-----GPGGP--GPDGQSGGCGN 232

RESULT 40  
 ID 017038 PRELIMINARY; PRT; 319 AA.  
 AC 017038;  
 DT 01-JUN-1998 (TEMBLrel. 05, Created)  
 DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
 DE Hypochemical 31.3 kDa protein.  
 GN T15B7.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Peioderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Pauley A., Gattung S.;  
 RT "The sequence of C. elegans coemid T15B7,";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission,";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF022885; AAB69959.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002486; Col\_cuticle\_N.  
 DR Pfam: PF01391; Collagen; 3.  
 DR Pfam: PF01484; Col\_cuticle\_N; 1.  
 DR Hypochemical protein.  
 KW SEQUENCE 319 AA; 31292 MW; 5242AAC668A8488 CRC64;

Query Match  
 Best Local Similarity 5.8%; Score 141; DB 5; Length 319;  
 Matches 90; Conservative 27; Mismatches 111; Indels 124; Gaps 23;

QY 13 VGLALLLGVAGVSGSL-----EPYVNSAKRFOAGSG-----VLYVPOIGDRDLILC 62  
 DB 1 MSASTLVTAASAGIAIVCVFTVGTGTFNDINSFYDEKIGELKEKYGQIA--WQAMT 58

QY 63 PRAPRPGHSSPNVEFKVLIVG-----GAQRRCER-PPAPNLLITCDRPLDL 111  
 DB 59 PTPRSSGSS-----FLLGKRRQABCNCEQSGRGCPAPGPP-----GQPG-- 101

QY 112 RFTIKPQFVSFNLTGHEFRSHHDYIIATSDGTREGLESLOGVCLTRGMKYLRLVSGSP 171  
 DB 102 ---ARGAGLPGIAGQ-----PGSGARINATGRPFCIT-----CP 135

QY 172 RCGAVPRKPEVSEMPMERDRGAASLEPKENLPDPTSNATSRGAEGPLPPPSMPAVAGA 231  
 DB 136 AGAPGAPGP-----PGA-----PGPXNNQGPAPAPQS--GGRGP--GPPRGP--AGD 177

QY 232 AGGLALLLLGVAGAGACWRRRAKPSBSRHPG-PGSPFRGG--SLGLGGGGG-----WG 284  
 DB 178 AG-----SHGQGHGHSFONPFRGQGRSRGTGASGRPGQG 214

QY 285 PREAPEBELGIALRGGAADP-PFCPHYEKVSGDYGHPIVVDGPPQSPPN 335  
 DB 215 PAGA-PGQPG--RSGGAGTPGPGP-----PGSGQPGHSGNDGVYGTPTGN 257

RESULT 41  
 ID 093485 PRELIMINARY; PRT; 809 AA.  
 AC 093485;  
 DT 01-NOV-1998 (TEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
 DE Alpha 1 type I collagen (Fragment) (Salmo gairdneri).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RTROBLAST;  
 RA Saito M., Kuniaki N., Hiroto I., Aoki T., Ishida M., Urano N.,  
 RA Kimura S.;  
 RT "Partial characterization of cDNA clones encoding the three distinct  
 RT pro alpha chains of type I collagen from rainbow trout.";  
 RL Fisheries Sci. 64:780-786(1998).  
 RN EMBL: AB008373; BAB33380.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 9.  
 DR Pfam: PF000007; Collagen; 1.  
 DR Prodom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 KW Collagen.  
 KW NON TER  
 FT SEQUENCE 809 AA; 78164 MW; 68C056A7640FCA81 CRC64;

Query Match  
 Best Local Similarity 5.8%; Score 141; DB 13; Length 809;  
 Matches 61; Conservative 15; Mismatches 93; Indels 54; Gaps 9;

QY 139 ATSDGTREGLESLOG-----GVCLTRGMKYLRLVSGSPRGAVPRKPEVSEMPMERDGA 193  
 DB 98 ADGAGGKDVGRGWTGPIGPNPAGSPGDK-----GETGAGAVGPGSARAPGERGSGA 152

QY 194 HSLPEKENVLPDPTSNAT-----SRGAEPLPPSPMPAVAGAGGLALLL 239  
 DB 153 ---PGAPGAPPGGGGQPGAGKEADNGAKDQAGQAGFTGAPGPGPAGN----- 203

QY 240 LGVAGAGACWRRRAKPSBSRHPG-----PGSPFRGSLGLGGGCGMPREAEFG 292  
 DB 204 TGAGAGAGA-----AGPPATGPGAAGRFGPBPGRNNGPPGTGPGGKEGQXGNRG 257

QY 293 LGIALRGG--GAADPPCFPHYEKVSGDYGHPIVVDGPPQSP 333

Db 258 TGPAGRPGEELGAGPP-----GPKGKGQPGGNDGPNGPSCTP 294

## RESULT 42

Q910C0

ID 0910C0 PRELIMINARY; PRT; 1449 AA.

AC 0910C0;

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, last annotation update)

DE Collagen a1(I).

GN COL1A1.

OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;

OC Proteocephaletergii; Salmoniformes; Salmonidae; Oncorhynchus.

OC NCBI\_TaxID=8022;

OX NCBI\_TaxID=8022;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21257802; PubMed=11358497;

RA Saito M., Takenouchi Y., Kunisaki N., Kimura S.,

RT "Complete primary structure of rainbow trout type I collagen

consisting of a1(I)2(I)3(I) heterotrimers."

RL Eur. J. Biochem. 268:2817-2827(2001).

DR EMBL; AB052835; BAB55661.1; -

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR000885; Fib collagen\_C.

DR InterPro; IPR001007; vWF\_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR Pfam; PF00093; vwc; 1.

DR ProDom; PD000007; Collagen; 2.

DR ProDom; PD002078; Fib collagen C; 1.

DR PROSITE; PS01208; vWFC; UNKNOWN\_1

SQ SEQUENCE 1449 AA; 137117 MW; 62BFB7A7BFD65288 CRC64;

## Query Match

Best Local Similarity 5.8%; Score 141; DB 13; Length 1449;

Matches 61; Conservative 15; Mismatches 93; Indels 54; Gaps 9;

QY 139 ATSTGTGEGLESLOG-----GVCITRGMKYLTVKVGSPRGAVPRKVSMPMERDRA 193

Db 738 ADGAGGAGDVGKMTGTPGPNGPSGSDK-----GEGTAGAVPGSARGAPEREBSGA 792

QY 194 HSLEPGKENTPGDPTSNAT-----SRGAEPLPPSPMAVAGAGLALL 239

Db 793 ----PGPAGRPPGSGGQPGAKGADNAGKDDGAGAGTGAAPGPGPAGN----- 843

QY 240 LGVAGAGGACWRRRRRAKPSERHPG-----PGSFGGSLGLGGGGGMPREAPGE 292

Db 844 TGAAGAGGA-----AGPGATGPPGAAGRFGPPGSGNNPGTGPGGKGGKGNRGE 897

QY 293 LGTALRGS--GAADPPCPHYEKVSGDYGHPIVVDGPGSP 333

Db 898 TGPAGRPGEELGAGPP-----GPKGKGQPGGNDGPNGPSCTP 934

## RESULT 43

Q90YCS

ID 090YCS PRELIMINARY; PRT; 219 AA.

AC 090YCS;

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, last sequence update)

DT 01-MAR-2002 (TREMblrel. 20, last annotation update)

DE Ephrin-A3.

GN EPHRIN-A3.

OS Brachydanio rerio (Zebrafish) (Zebra danio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OC NCBI\_TaxID=7955;

OX NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21412237; PubMed=11520665;

RA Hirata Y., Mada M., Harada T., Yamasu K., Okamoto H.,

RT "Identification of ephrin-A3 and novel genes specific to the midbrain-

MB in embryonic zebrafish by ordered differential display."

RL Mech. Dev. 107:83-96(2001).

DR EMBL; AB051678; BAB55991.1; -

DR InterPro; IPR001799; Ephrin.

DR Pfam; PF00812; Ephrin; 1.

DR ProDom; PD002533; Ephrin; 1.

DR PROSITE; PS01299; EPHRIN; UNKNOWN\_1.

SQ SEQUENCE 219 AA; 25146 MW; 7151927B03F8EA01 CRC64;

## Query Match

Best Local Similarity 5.7%; Score 140.5; DB 13; Length 219;

Matches 58; Conservative 27; Mismatches 92; Indels 39; Gaps 9;

QY 33 VYNSANKRFQAEQGVYVYPOIGRIDLDCPRAPPGHSSPN--YEYKLYLVGAQGR 90

Db 28 VHMSSNILLRKE-CYTLQVAVNDYLDICPHY-----NSSQRGIAEGYLVWVSYNGYR 81

QY 91 REAPAPVNLITCDR---PDLDRFTTKFQEQYSPNLMGHEFRSHHDYITATSDGTREG 147

Db 82 TCD-PQLGFKWECNRPAPAPAPKTFSEKFORYSAFSLGYEFHYGOEYYISTP----- 134

QY 148 LESLGGYCLTRGMKYLTVKVGSPRGAVPRKVSMPMERDRAHSLPGKENTLPDGP 207

Db 135 -THHGRSCRLRRTVYCCSTASDSDDEPQTP-----DYTLRP--NIKIDD 178

QY 208 TSNATSRGAEGLPPSPMAVAGAGLALLGVA 243

Db 179 LDD-----YDNPEVYKLEKISGSSPSRDLTLTVA 209

## RESULT 44

P91274

ID P91274 PRELIMINARY; PRT; 305 AA.

AC P91274;

DT 01-MAY-1997 (TREMblrel. 03, Created)

DT 01-MAY-1997 (TREMblrel. 03, last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)

DE Hypochemical 30.8 kDa protein.

GN P26B1.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Pelodidae; Caenorhabditis.

OC NCBI\_TaxID=6239;

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium."

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA latreille P., Wamsley P., Kramer J.;

RT "The sequence of C. elegans cosmid F26B1."

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waseron R.;

RT "Direct Submission."

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; U08444; AAB37788.1; -

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR002486; Col cuticle\_N.

DR Pfam; PF01484; Col cuticle\_N; 1.

DR Hypothetical protein.

KM

SQ SEQUENCE 305 AA; 30770 MW; 6BDC3ADD3E90A358 CRC64;





## RESULT 47

Q9YIB4 PRELIMINARY; PRT; 1450 AA.  
 ID Q9YIB4  
 AC Q9YIB4  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Alpha 1 type I collagen.  
 OS Cyrops pyrrhogaster (Japanese common newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Caudata; Salamandridae; Cynops.  
 OX NCBI\_Taxid=8330;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=REGENERATE FORELIMBS;  
 RX MEDLINE=99407244; PubMed=10474166;  
 RA Asahina K., Obata M., Yoshizato K.;  
 RT "Expression of genes of type I and type II collagen in the formation  
 and development of the blastema of regenerating newt limb.";  
 RL Dev. Dyn. 216:59-71(1999).  
 DR EMBL; AB015438; BAA36973.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR Pfam; PF01410; COLF1; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR ProDom; PD000007; Collagen; 1.  
 DR ProDom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLF1; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; VWC; UNKNOWN\_1.  
 KM Collagen.  
 SQ SEQUENCE 1450 AA; 137564 MW; ABF8A74841B87BC CRC64;

Query Match 5.7%; Score 139.5; DB 13; Length 1450;  
 Best Local Similarity 25.7%; Pred. No. 0.023;  
 Matches 103; Conservative 24; Mismatches 137; Indels 137; Gaps 24;

QY 2 GPHSGPGVAVGALLGLVGL--VSGLSPEPYWNSANKRFGAGGYLVPGIGRDL 59  
 DB 552 GPP--GPGAGAGGQ---GVVGFPGPKGAGP-----GKSGER-- 584  
 QY 60 LCPARPDPGSHSPNVEFYLYVGAQGRCAPAPNLLTCDSPDIDRTTIFQ 119  
 DB 585 ---GVAGPFGATGAPGKXG---EAGQG---PGPS-GPGERGEGPAGSPFGQ 630  
 QY 120 Y--SPNLGHPFRSHHDYIIATSDGTREGLSLQG--GVCLTRGMKVLNVQSPRGA 175  
 DB 631 LPGDFPAGEGAKPGEQ--APGDAGGPGSPGPRGERGPPGERG-----GGPFAAQ 680  
 QY 176 VPR---KPVSE-----MPVER-----DRGAHSLPEP 199  
 DB 681 GPRGSPGPGNDGAKGKAGAPGCRPGPLQGMPPERGSGAPGAKGGRGA-----G 725  
 QY 200 KENLPDPTNATRGAGAPLPSPSMRAVAG--AAGIALILLGVAAGAGMCRRRRAK 257  
 DB 736 TKGADGAGKXG--ARGLPIGPFGPAGAPADKKEGGPS---GPAAPTGA-----RGS 784  
 QY 258 PSESRHPG-----PGSPRGSGSLG---LGGGGMGMPREAPGLGIAL 297  
 DB 785 PGERGERPAPAPAGICGPPGADGPGAKGSGDAGPPKADGAPAPAGPTGA--PGAGNVG 843  
 QY 298 RGG---GAADPPCPHYEKVSGDYGHVYITVGDGPQSP 333  
 DB 844 APGPKTGAGAGPPAGATGFPGAAGRLGPPGSGNAGAPPPPP 884

## RESULT 48

Q9UMG6 PRELIMINARY; PRT; 1690 AA.  
 ID Q9UMG6  
 AC Q9UMG6;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Collagen type IV a6 chain.  
 GN COL4A6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE OF 1-1644 FROM N.A.  
 RX MEDLINE=9629642; PubMed=866106;  
 RA Zhang X., Zhou J., Redders S.T., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
 in Alport syndrome-associated telomylomatosis";  
 RL Genomics 33:473-479(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zhang X.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U47004; AAB19039.1; -  
 DR EMBL; U46980; AAB19039.1; JOINED.  
 DR EMBL; U46961; AAB19039.1; JOINED.  
 DR EMBL; U46962; AAB19039.1; JOINED.  
 DR EMBL; U46963; AAB19039.1; JOINED.  
 DR EMBL; U46964; AAB19039.1; JOINED.  
 DR EMBL; U46965; AAB19039.1; JOINED.  
 DR EMBL; U46966; AAB19039.1; JOINED.  
 DR EMBL; U46967; AAB19039.1; JOINED.  
 DR EMBL; U46968; AAB19039.1; JOINED.  
 DR EMBL; U46969; AAB19039.1; JOINED.  
 DR EMBL; U46970; AAB19039.1; JOINED.  
 DR EMBL; U46971; AAB19039.1; JOINED.  
 DR EMBL; U46972; AAB19039.1; JOINED.  
 DR EMBL; U46973; AAB19039.1; JOINED.  
 DR EMBL; U46974; AAB19039.1; JOINED.  
 DR EMBL; U46975; AAB19039.1; JOINED.  
 DR EMBL; U46976; AAB19039.1; JOINED.  
 DR EMBL; U46977; AAB19039.1; JOINED.  
 DR EMBL; U46978; AAB19039.1; JOINED.  
 DR EMBL; U46979; AAB19039.1; JOINED.  
 DR EMBL; U46980; AAB19039.1; JOINED.  
 DR EMBL; U46981; AAB19039.1; JOINED.  
 DR EMBL; U46982; AAB19039.1; JOINED.  
 DR EMBL; U46983; AAB19039.1; JOINED.  
 DR EMBL; U46984; AAB19039.1; JOINED.  
 DR EMBL; U46985; AAB19039.1; JOINED.  
 DR EMBL; U46986; AAB19039.1; JOINED.  
 DR EMBL; U46987; AAB19039.1; JOINED.  
 DR EMBL; U46988; AAB19039.1; JOINED.  
 DR EMBL; U46989; AAB19039.1; JOINED.  
 DR EMBL; U46990; AAB19039.1; JOINED.  
 DR EMBL; U46991; AAB19039.1; JOINED.  
 DR EMBL; U46992; AAB19039.1; JOINED.  
 DR EMBL; U46993; AAB19039.1; JOINED.  
 DR EMBL; U46994; AAB19039.1; JOINED.  
 DR EMBL; U46995; AAB19039.1; JOINED.  
 DR EMBL; U46996; AAB19039.1; JOINED.  
 DR EMBL; U46997; AAB19039.1; JOINED.  
 DR EMBL; U46998; AAB19039.1; JOINED.  
 DR EMBL; U46999; AAB19039.1; JOINED.  
 DR EMBL; U47000; AAB19039.1; JOINED.  
 DR EMBL; U47001; AAB19039.1; JOINED.  
 DR EMBL; U47002; AAB19039.1; JOINED.  
 DR EMBL; U47003; AAB19039.1; JOINED.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD000007; Collagen; 23.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KM Collagen.  
 SQ SEQUENCE 1690 AA; 163696 MW; 4698AE2CC2D3E859 CRC64;

Query Match 5.7%; Score 139.5; DB 4; Length 1690;  
 Best Local Similarity 25.2%; Pred. No. 0.027;  
 Matches 61; Conservative 25; Mismatches 83; Indels 73; Gaps 11;

QY 147 GLESLQG--GVCLTRGMKVL-----RVGQSPRGA-----VPRKPVSM 184  
 DB 905 GFGPIRPLPGLISGRGLKGI:PGSTGKMGPGGRGTPGEKDRGNPQVGI:PSRRRMSNL 964  
 QY 185 PMERDRGAHSL-----EPKENTLPDPTSNATSRGAG-PLPP----- 222  
 DB 965 WLKGDKSGSQSAGSNPGRPGRDKGAGRP:PPGLPGLPGLPGLI:KGVSGKPGPFVGI 1024  
 QY 223 -----PSMPVAGAAG-----GL--ALLLGVAGAGGACWRRRAKXSE 260  
 DB 1025 RGLPGLKSSGIGTFPGMPESGSGIRGSPGLPGASGLPGLKGDNGQTV--EISGSPG 1082  
 QY 261 SRHPGPGSF-GRGSLGLGGGGMGPPEAPPELGI:ALRGGAADPPFCPHYEKVSGDYG 319  
 DB 1083 KQPGSGSGFGTGRDGLIGNIGFPGKKGEDKGVGS----GDVGLPGA:PGFVAGMRG 1138  
 QY 320 HP 321  
 DB 1139 EP 1140

RESULT 49  
 O9Y4L4 PRELIMINARY; PRT: 1691 AA.

ID O9Y4L4  
 AC O9Y4L4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Collagen type IV a6 chain.  
 GN COL4A6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-1645 FROM N.A.  
 RX MEDLINE=96299642; PubMed=8661006;  
 RA Zhang X., Zhou J., Reeder S.T., Trygsvaen K.;  
 RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
 in Alport syndrome-associated leiomyomatosis.";  
 RL Genomics 33:473-479(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zhang X.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; U46982; AAB19038.1; JOINED.  
 DR EMBL; U46983; AAB19038.1; JOINED.  
 DR EMBL; U46984; AAB19038.1; JOINED.  
 DR EMBL; U46985; AAB19038.1; JOINED.  
 DR EMBL; U46986; AAB19038.1; JOINED.  
 DR EMBL; U46987; AAB19038.1; JOINED.  
 DR EMBL; U46988; AAB19038.1; JOINED.  
 DR EMBL; U46989; AAB19038.1; JOINED.  
 DR EMBL; U46990; AAB19038.1; JOINED.  
 DR EMBL; U46991; AAB19038.1; JOINED.  
 DR EMBL; U46992; AAB19038.1; JOINED.  
 DR EMBL; U46993; AAB19038.1; JOINED.  
 DR EMBL; U46994; AAB19038.1; JOINED.  
 DR EMBL; U46995; AAB19038.1; JOINED.  
 DR EMBL; U46996; AAB19038.1; JOINED.  
 DR EMBL; U46997; AAB19038.1; JOINED.  
 DR EMBL; U46998; AAB19038.1; JOINED.  
 DR EMBL; U46999; AAB19038.1; JOINED.  
 DR EMBL; U47000; AAB19038.1; JOINED.  
 DR EMBL; U47001; AAB19038.1; JOINED.  
 DR EMBL; U47002; AAB19038.1; JOINED.  
 DR EMBL; U47003; AAB19038.1; JOINED.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 23.  
 DR ProDom; PD000007; Collagen; 4.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1691 AA; 163873 MW; 769AA53DD1C7CA87 CRC64;

DR EMBL; U46982; AAB19038.1; JOINED.  
 DR EMBL; U46983; AAB19038.1; JOINED.  
 DR EMBL; U46984; AAB19038.1; JOINED.  
 DR EMBL; U46985; AAB19038.1; JOINED.  
 DR EMBL; U46986; AAB19038.1; JOINED.  
 DR EMBL; U46987; AAB19038.1; JOINED.  
 DR EMBL; U46988; AAB19038.1; JOINED.  
 DR EMBL; U46989; AAB19038.1; JOINED.  
 DR EMBL; U46990; AAB19038.1; JOINED.  
 DR EMBL; U46991; AAB19038.1; JOINED.  
 DR EMBL; U46992; AAB19038.1; JOINED.  
 DR EMBL; U46993; AAB19038.1; JOINED.  
 DR EMBL; U46994; AAB19038.1; JOINED.  
 DR EMBL; U46995; AAB19038.1; JOINED.  
 DR EMBL; U46996; AAB19038.1; JOINED.  
 DR EMBL; U46997; AAB19038.1; JOINED.  
 DR EMBL; U46998; AAB19038.1; JOINED.  
 DR EMBL; U46999; AAB19038.1; JOINED.  
 DR EMBL; U47000; AAB19038.1; JOINED.  
 DR EMBL; U47001; AAB19038.1; JOINED.  
 DR EMBL; U47002; AAB19038.1; JOINED.  
 DR EMBL; U47003; AAB19038.1; JOINED.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001442; ProcollagenC4.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 23.  
 DR ProDom; PD000007; Collagen; 4.  
 DR ProDom; PD003923; ProcollagenC4; 1.  
 DR SMART; SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1691 AA; 163873 MW; 769AA53DD1C7CA87 CRC64;

Query Match 5.7%; Score 139.5; DB 4; Length 1691;  
 Best Local Similarity 25.2%; Pred. No. 0.027;  
 Matches 61; Conservative 25; Mismatches 83; Indels 73; Gaps 11;

QY 147 GLESLQG--GVCLTRGMKVL-----RVGQSPRGA-----VPRKPVSM 184  
 DB 906 GFGPIRPLPGLISGRGLKGI:PGSTGKMGPGGRGTPGEKDRGNPQVGI:PSRRRMSNL 965  
 QY 185 PMERDRGAHSL-----EPKENTLPDPTSNATSRGAG-PLPP----- 222  
 DB 966 WLKGDKSGSQSAGSNPGRPGRDKGAGRP:PPGLPGLPGLPGLI:KGVSGKPGPFVGI 1025  
 QY 223 -----PSMPVAGAAG-----GL--ALLLGVAGAGGACWRRRAKXSE 260  
 DB 1026 RGLPGLKSSGIGTFPGMPESGSGIRGSPGLPGASGLPGLKGDNGQTV--EISGSPG 1083  
 QY 261 SRHPGPGSF-GRGSLGLGGGGMGPPEAPPELGI:ALRGGAADPPFCPHYEKVSGDYG 319  
 DB 1084 KQPGSGSGFGTGRDGLIGNIGFPGKKGEDKGVGS----GDVGLPGA:PGFVAGMRG 1139  
 QY 320 HP 321  
 DB 1140 EP 1141

RESULT 50  
 O9TEU5 PRELIMINARY; PRT: 705 AA.

ID O9TEU5  
 AC O9TEU5;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE FLJ00201 protein (Fragment).  
 GN FLJ00201.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN;

RA Jikuya H., Takano J., Nomura N., Kiyuno R., Nagase T., Ohara O.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 spleen.";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK074129; BAB84955.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 705 AA; 67430 MM; 07DB85A65A948ED3 CRC64;

Query Match 5.7%; Score 139; DB 4; Length 705;  
 Best Local Similarity 25.8%; Pred. No. 0.011;  
 Matches 73; Conservative 18; Mismatches 88; Indels 104; Gaps 14;

```

QY 147 GLESLQG-----GYCLTRGKMKVLLRWGQSPRGAVP-----RKPYSEMPMERDRGAA 193
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 289 GLPGFQGSBSGAKGERGTGPPGLI-----GPTGYGMFGILPGPKDRGPAGVPGILGDRG-- 342
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 194 HSLPCKENLPD-----PTSNAT-----SRGAEGLPLPPSPMPAVAGAAG 233
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 343 --EPGDEDEPGEQPGQIGPPGLPGSAGLPGRGRPPGPKGEAGPGGPPGPGIRGDG 399
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 234 --GLALLLVAGAGAGAMCWRRRRAKPSRSRHPG-----GSFGRG 272
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 400 PSGLA-----GRGVPG-----ERGLPGAGPPPGTGPKEGPFTGRPGGPGVAGALGOK 449
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 273 GSLGI-----GGGGMGPR-----EAPGELGIALRG-----GGAAD 304
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 450 GDLGLPGQPGRLRGPESGIPGLQPGAPFIPGQGLPGKKEPGLPGPPGGGRAGEPETAAGPTG 509
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 305 PPFCHYKVKVSGDYGHFPIYIVQDGPQSPFNITYTSSVLEMP 347
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 510 PPGVPGSPGITGPPGP--GPPGPPGAPGAFDETGIAGLHLP 549
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Search completed: February 11, 2003, 12:04:54  
 Job time : 49.7673 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 11:47:24 ; Search time 19.4591 Seconds

(without alignments)  
969.814 Million cell updates/sec

Title: US-10-021-121-2

Perfect score: 2450  
Sequence: 1 MGPPIHSGRGVGVNALLLG.....TTLLRGRASVNAEGQHGPL 455

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 50 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	184.1	75.1	340	EPB3_HUMAN	Q15766 homo sapien
2	177.1	72.3	340	EPB3_MOUSE	O35393 mus musculu
3	628.5	25.7	336	EPB2_MOUSE	P52800 mus musculu
4	623.4	25.4	346	EPB1_HUMAN	P98172 homo sapien
5	620.5	25.3	333	EPB2_HUMAN	P52799 homo sapien
6	619.5	25.3	334	EPB1_CHICK	O73612 gallus gall
7	617.5	25.2	332	EPB2_HUMAN	O73874 brachydant
8	604.5	24.7	345	EPB1_MOUSE	P52795 mus musculu
9	599.5	24.5	345	EPB1_RAT	P52796 ratu mus
10	591.5	24.1	327	EPB1_XENLA	O13097 xenopus lae
11	185.5	7.6	195	EPB2_HUMAN	P73727 brachydant
12	179.5	7.3	238	EPB3_HUMAN	P52797 homo sapien
13	176.5	7.2	209	EPB2_MOUSE	P52801 mus musculu
14	175.5	7.2	213	EPB2_HUMAN	O43921 homo sapien
15	172.5	7.0	200	EPB2_CHICK	P52802 gallus gall
16	170.5	6.9	228	EPB5_CHICK	P52804 gallus gall
17	169.5	6.9	228	EPB4_HUMAN	P52798 homo sapien
18	169.5	6.9	228	EPB5_HUMAN	P73728 brachydant
19	167.5	6.8	216	EPB1_XENLA	P52794 xenopus lae
20	167.5	6.8	228	EPB5_HUMAN	P52803 homo sapien
21	167.5	6.8	228	EPB5_MOUSE	O08543 mus musculu
22	167.5	6.8	228	EPB5_RAT	P97605 ratu mus
23	166.5	6.8	205	EPB1_HUMAN	P73727 brachydant
24	162.5	6.6	205	EPB1_RAT	P73727 brachydant
25	161.5	6.6	205	EPB1_MOUSE	P73727 brachydant
26	160.5	6.6	206	EPB4_MOUSE	O08542 mus musculu
27	159.5	6.5	680	CA1A_MOUSE	O08543 mus musculu
28	154.5	6.3	1049	CA1B_BOVIN	P04258 bos tauru
29	151.5	6.2	301	CCO2_CAEEL	P17656 caenorhadi
30	148.5	6.1	1670	CA14_HUMAN	Q01955 homo sapien
31	146.5	6.0	1527	CA1H_MOUSE	P30754 rificia pach
32	146.5	6.0	1027	CA1F_RIFPA	P23206 bos tauru
33	145.5	5.9	674	CA1A_BOVIN	P23206 bos tauru

34	144.5	5.9	1516	1	CA1H_HUMAN	P39060 homo sapien
35	144	5.9	675	1	CA19_CHICK	P32017 gallus gall
36	144	5.9	921	1	CA19_HUMAN	P20849 homo sapien
37	143	5.8	674	1	CA1A_CHICK	P08125 gallus gall
38	142.5	5.8	635	1	CA28_HUMAN	P25067 homo sapien
39	142.5	5.8	1496	1	CA25_HUMAN	P05997 homo sapien
40	142	5.8	744	1	CA18_HUMAN	P27658 homo sapien
41	142	5.8	1029	1	CA26_MOUSE	Q02788 mus musculu
42	142	5.8	1763	1	CA24_ASCSU	P27393 ascaris suu
43	141.5	5.8	1466	1	CA13_HUMAN	P02461 homo sapien
44	141	5.8	744	1	CA18_RABIT	P14282 oryctolagus
45	140.5	5.7	743	1	CA18_MOUSE	Q00780 mus musculu
46	140	5.7	680	1	CA1A_HUMAN	Q01692 homo sapien
47	139.5	5.7	220	1	CA2C_HAECO	P16252 haemonchus
48	139	5.7	1804	1	CA1B_MOUSE	Q61245 mus musculu
49	138.5	5.7	1459	1	CA12_MOUSE	P28481 mus musculu
50	138.5	5.7	1736	1	CA2B_HUMAN	P13642 homo sapien

## ALIGNMENTS

RESULT 1  
ID EPB3\_HUMAN STANDARD; PRT; 340 AA.  
AC Q15766; 000680; Q92875;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE EPHrin-B3 precursor (EPH-related receptor tyrosine kinase ligand 8)  
DE (LEK-8) (EPH-related receptor transmembrane ligand Etk-L3).  
GN EPB3 OR EPLG8 OR LERK8.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cerretti D.P.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97271551; PubMed=9126477;  
RX Tang X.X., Pleasure D.E., Ikegaki N.;  
RT "CDNA cloning, chromosomal localization, and expression pattern of  
RT EPLG8, a new member of the EPLG gene family encoding ligands of EPH-  
RT related protein-tyrosine kinase receptors.";  
RL Genomics 41:17-24(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=96404527; PubMed=8808709;  
RX Gale N.W., Platenik A., Compton D.C., Jenkins N.A., Copeland N.G.,  
RA Gilbert D.J., Davis S., Wilkerson D.G., Yancopoulos G.D.,  
RT "Etk-L3, a novel transmembrane ligand for the Eph family of receptor  
RT tyrosine kinases, expressed in embryonic floor plate, roof plate and  
RT hindbrain segments.";  
RL Oncogene 13:1343-1352(1996).  
CC - FUNCTION: MAY PLAY A PIVOTAL ROLE IN FOREBRAIN FUNCTION. BINDS TO,  
CC AND INDUCE THE COLLAPSE OF COMMISSURAL AXONS/GROWTH CONES IN  
CC VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF  
CC LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).  
CC - SUBCELLULAR LOCATION: Type I membrane protein.  
CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN; EXPRESSED IN  
CC EMBRYONIC FLOOR PLATE, ROOF PLATE AND HINDBRNIN SEGMENTS.  
CC - SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U57001; AAB05170.1; -

DR EMBL; U66406; AAC51203.1; -

DR EMBL; U62775; AAC50707.1; -

DR Genew; HGNC:3228; EFNB3.

DR MIM; 602297; -

DR InterPro; IPR01799; Ephrin.

DR Pfam; PF00812; Ephrin; 1.

DR PRINTS; PR01347; EPHRIN.

DR Prodom; PD002533; Ephrin; 1.

DR PROSITE; PS01299; EPHRIN; 1.

KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;

KM Signal; Polymorphism.

FT SIGNAL 1 27

FT CHAIN 28 340

FT DOMAIN 28 226

FT TRANSMEM 227 247

FT DOMAIN 248 340

FT DOMAIN 338 340

FT CARBOHYD 210 210

FT VARIANT 166 166

FT R -> O.

FT FTID=VAR\_002356.

SO SEQUENCE 340 AA; 35834 MW; EDFF2A23C2DE79F CRC64;

Query Match 75.1%; Score 1841; DB 1; Length 340;

Best Local Similarity 100.0%; Pred. No. 8.6e-115;

Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPHSGGGVGVGALLLLGLVLSGLSEPYWMSANKRFQAEGGVYLPQIGDRIDL 60

DB 1 MGPHSGGGVGVGALLLLGLVLSGLSEPYWMSANKRFQAEGGVYLPQIGDRIDL 60

QY 61 LCPRARPPGPHSSPNVEYKYLYVGAQGRCEAPAPNLLTCDRDLDRFTIKQOEY 120

DB 61 LCPRARPPGPHSSPNVEYKYLYVGAQGRCEAPAPNLLTCDRDLDRFTIKQOEY 120

QY 121 SPNLMGHEFRSHHDYIATSDGTREGLESIGGVCLTRGKVLARVQSGPRGAVPRKP 180

DB 121 SPNLMGHEFRSHHDYIATSDGTREGLESIGGVCLTRGKVLARVQSGPRGAVPRKP 180

QY 181 VSEPMERDRGAASLEPGKENTPGDPTSNATSRGAGPPLPPSPMPAVAGAAGLALLLL 240

DB 181 VSEPMERDRGAASLEPGKENTPGDPTSNATSRGAGPPLPPSPMPAVAGAAGLALLLL 240

QY 241 GVAAGAGMCMRRRRARPSERHAPGSGFGRGSLGIGGGGCMGPRAPAECEGIALRG 300

DB 241 GVAAGAGMCMRRRRARPSERHAPGSGFGRGSLGIGGGGCMGPRAPAECEGIALRG 300

QY 301 GAADPFCEPHYKESGDIYVODGPPOSPPNITY 338

DB 301 GAADPFCEPHYKESGDIYVODGPPOSPPNITY 338

RESULT 2

EFNB3 MOUSE STANDARD; PRT; 340 AA.

ID EFNB3 MOUSE

AC 035393;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Ephrin-B3 precursor.

GN EFNB3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SSOURCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=98143367; PubMed=9484836;

RA Bergemann A.D., Zhang L., Chiang M.-K., Brambilla R., Klein R.,

RA Planagan J.G.;

RT "Ephrin-B3, a ligand for the receptor EphB3, expressed at the midline of the developing neural tube.";

RL Oncogene 16:471-480(1998).

RN [2]

RP FUNCTION.

RX MEDLINE=20171264; PubMed=10704386;

RA Imordt R., Wideman C., Kaprielian Z.;

RT "Complementary expression of transmembrane ephrins and their receptors in the mouse spinal cord: a possible role in constraining the orientation of longitudinally projecting axons.";

RL Development 127:1397-1410(2000).

CC -1- FUNCTION: MAY PLAY A PIVOTAL ROLE IN FOREBRAIN FUNCTION. BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF LONGITUDINALLY PROJECTING AXONS.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS, SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE PERIOD OF COMMISSURAL AXON PATHFINDING.

CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF025288; AAC53537.1; -

DR MGD; MGI:109196; Efmb3.

DR InterPro; IPR001799; Ephrin.

DR Pfam; PF00812; Ephrin; 1.

DR PRINTS; PR01347; EPHRIN.

DR Prodom; PD002533; Ephrin; 1.

DR PROSITE; PS01299; EPHRIN; 1.

KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;

KM Signal.

FT SIGNAL 1 27

FT CHAIN 28 340

FT DOMAIN 28 227

FT TRANSMEM 228 248

FT DOMAIN 249 340

FT DOMAIN 338 340

FT CARBOHYD 210 210

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 340 AA; 35884 MW; 52F3D58FDD09A6B8 CRC64;

Query Match 72.3%; Score 1771; DB 1; Length 340;

Best Local Similarity 95.6%; Pred. No. 3.6e-110;

Matches 323; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 MGPHSGGGVGVGALLLLGLVLSGLSEPYWMSANKRFQAEGGVYLPQIGDRIDL 60

DB 1 MGPHSGGGVGVGALLLLGLVLSGLSEPYWMSANKRFQAEGGVYLPQIGDRIDL 60

QY 61 LCPRARPPGPHSSPNVEYKYLYVGAQGRCEAPAPNLLTCDRDLDRFTIKQOEY 120

DB 61 LCPRARPPGPHSSPNVEYKYLYVGAQGRCEAPAPNLLTCDRDLDRFTIKQOEY 120

QY 121 SPNLMGHEFRSHHDYIATSDGTREGLESIGGVCLTRGKVLARVQSGPRGAVPRKP 180

DB 121 SPNLMGHEFRSHHDYIATSDGTREGLESIGGVCLTRGKVLARVQSGPRGAVPRKP 180

QY 181 VSEPMERDRGAASLEPGKENTPGDPTSNATSRGAGPPLPPSPMPAVAGAAGLALLLL 240

DB 181 VSEPMERDRGAASLEPGKENTPGDPTSNATSRGAGPPLPPSPMPAVAGAAGLALLLL 240

QY 241 GVAAGAGMCMRRRRARPSERHAPGSGFGRGSLGIGGGGCMGPRAPAECEGIALRG 300

DB 241 GVAAGAGMCMRRRRARPSERHAPGSGFGRGSLGIGGGGCMGPRAPAECEGIALRG 300

```

DB      241 GAGAGACACWRRARRRPSSESRHPPGSGFGRGSLGLGGCGGNGPREAREPGEIAGLRG 300
QY      301 GADPFCEPHYKSGDYGHPIYIVODGPPQSPNNIY 338
DB      301 GTADPFCEPHYKSGDYGHPIYIVODGPPQSPNNIY 338

RESULT 3
ID      EFB2_MOUSE      STANDARD;      PRT;      336 AA.
AC      P52800;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)
DE      (LERK-5) (HTK ligand) (ELF-2).
GN      EFN2 OR EPLG5 OR LERKS OR HTKL OR ELF2 OR EPL5.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX      NCBI_TaxID=10090;
[1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96145238; PubMed=8559144;
RA      Carrecci D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,
RA      Maraskovsky E., Park L.S., Lyman S.D., Copeland N.G.,
RA      Gilbert D.O., Jenkins N.A., Fletcher R.A.;
RT      "Isolation of LERK-5: a ligand of the eph-related receptor tyrosine
RT      kinases."
RT      Mol. Immunol. 32:1197-1205(1995).
[2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=CB57BL/6J X SJL/J;
RX      MEDLINE=9519254; PubMed=7534404;
RA      Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D.,
RA      Gillett N., Matthews W.;
RT      "Molecular cloning of a ligand for the EPH-related receptor protein-
RT      tyrosine kinase Htk."
RT      Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).
[3]
RP      SEQUENCE FROM N.A.
RX      STRAIN=ICR; TISSUE=Brain;
RX      MEDLINE=95319837; PubMed=7651410;
RA      Beremann A.D., Cheng H.J., Brambilla R., Klein R., Flanagan J.G.;
RA      "ELF-2, a new member of the Eph ligand family, is segmentally
RT      expressed in mouse embryos in the region of the hindbrain and newly
RT      forming somites."
RT      Mol. Cell. Biol. 15:4921-4929(1995).
[4]
RP      FUNCTION.
RX      MEDLINE=20171264; PubMed=10704386;
RA      Imondi R., Wideman C., Kaprielian Z.;
RT      "Complementary expression of transmembrane ephrins and their receptors
RT      in the mouse spinal cord: a possible role in constraining the
RT      orientation of longitudinally projecting axons."
RT      Development 127:1397-1410(2000).
CC      -1- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
CC      LONGITUDINALLY PROJECTING AXONS.
CC      -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS,
CC      SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH
CC      THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL
CC      LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.
CC      -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE
CC      PERIOD OF COMMISSURAL AXON PATHFINDING.
CC      -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC      CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its

```

```

CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U16819; AAA97908.1; -
DR      EMBL; L36847; AAC42052.1; -
DR      EMBL; U30244; AAA82934.1; -
DR      MGD; MGI:105097; Efnb2.
DR      InterPro: IPR001799; Ephrin.
DR      Pfam: PF00812; Ephrin.1.
DR      PRINTS; PR01347; EPHRIN.
DR      PRODOM; PD002533; Ephrin.1.
DR      PROSITE; PS01299; EPHRIN; 1.
KM      Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KM      Signal; Phosphorylation.
FT      SIGNAL      1      28
FT      CHAIN      29      336
FT      DOMAIN      23      232
FT      TRANSMEM      233      253
FT      DOMAIN      254      336
FT      DOMAIN      334      336
FT      CARBOHYD      39      39
FT      CARBOHYD      142      142
FT      CONFLICT      3      4
FT      CONFLICT      177      177
SQ      SEQUENCE      336 AA; 37202 MW; D0869496B39554 CRC64;

Query Match      25.7%; Score 628.5; DB 1; Length 336;
Best Local Similarity 41.8%; Pred. No. 8, 8e-35;
Matches 137; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

QY      14 GALLILGVTLGVSGLSLPEYNNNSANKRFQAEQGVLYLPQIGRLDLCRAAPPGPHSS 73
DB      17 GLMLVLCRTAISRSIVLPEIYNNSSNKFPGQGLVLYPQIGDKLDICPKV--DSKTV 73
QY      74 PNYEFTLYLVGAQGRCAAPAPRNILLCDPRDDLTARTTFCQSPYPLMGHERSRH 133
DB      74 GQYEVYVYVMDQDQDCRTYKENTPILNCAAPDDVKTFTFOESPLMLWLEFOK 133
QY      134 DYIILATSDGTREGLESLOGVCLTRGMKVLRLVGQ--SPRGAVPRKPYSEPMER-DR 190
DB      134 DYIILSTNSGLELDQEGVQCTRAKMLKMGVDASAGSARHNGFRRDELEAGTN 193
QY      191 GAHSLSRPGKENTPGDTSNATSRGAGPLPPSRMAYVAGAGLLALLGVAGAGAMC 250
DB      194 GRSTTSPEVKPSPGSDTDSNAGSHGNNLGSFVALFAGIAGGCTIFVITLVVLL 253
QY      251 WRRRAKPSRRHPPGSPGFRGSLGLGGCGGNGPREAREPGEIAGLRGGAADPFCEPH 310
DB      254 KYRRRHAKHSPQHTTLLSLTATPKRGNN---NGSEPSDVITFLR---TADSVFCFH 306
QY      311 YEKVSGDYGHPIYIVODGPPQSPNNIY 338
DB      307 YEKVSGDYGHPIYIVODGPPQSPNNIY 334

RESULT 4
ID      EFB1_HUMAN      STANDARD;      PRT;      346 AA.
AC      P98172;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)
DE      (LERK-2) (ELK ligand) (ELK-L).
GN      EFN1 OR EPLG2 OR LERK2 OR EFL-3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
[1]
RP      SEQUENCE FROM N.A.

```

RC TISSUE=Placenta; PubMed=8070404;  
 RX MEDLINE=94349923; PubMed=8070404;  
 RA Beckmann M.P., Cerretti D.P., Baum P., Vandenberg T., James L.,  
 RA Farrah T., Kozlosky C., Hollingsworth T., Shilling H., Maraskovsky E.,  
 RA Fletcher F.A., Lhotak V., Pawson T., Lyman S.D.;  
 RT "Molecular characterization of a family of ligands for eph-related  
 RT tyrosine kinase receptors";  
 RL EMBO J. 13:3757-3762(1994).  
 RN [2]  
 RA SEQUENCE FROM N.A.  
 RA Davis S., Gale N.W., Aldrich T.H., Maisompierre P.C., Lhotak V.,  
 RA Pawson T., Goldfarb M., Yancopoulos G.D.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RA SEQUENCE FROM N.A.  
 RA Fletcher F.A., Huebner K., Shaffer L.G., Monaco A., Mueller U.,  
 RA Kozlosky C., Druck T., Simoneaux D.K., Fairweather N., Chelly J.,  
 RA Cerretti D.P., Belmont J.W., Beckmann M.P., Lyman S.D.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RA SEQUENCE FROM N.A.  
 RA Howden P.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL  
 CC AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE  
 CC ORIENTATION OF LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB1 AND EPHA1.  
 CC -1- BINDS GRP1 AND GRP2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: HEART, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE,  
 CC KIDNEY, PANCREAS.  
 CC -1- INDUCTION: BY TNF-ALPHA.  
 CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE  
 CC CYTOPLASMIC DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U09304; AAA53093.1; -  
 DR EMBL; L37361; AAA52369.1; -  
 DR EMBL; U09303; AAB41127.1; -  
 DR EMBL; A1136092; CAB6409.1; -  
 DR Genew; HANC3226; EPNBL.  
 DR MIM; 300035; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR PRINTS; PRO1347; EPHRIN.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KM Developmental protein; Neutrogenesis; Transmembrane; Glycoprotein;  
 KM Signal; Phosphorylation.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 346 EPHRIN-B1.  
 FT DOMAIN 25 237 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT DOMAIN 259 346 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 344 346 PDZ RECOGNITION MOTIF (POTENTIAL).  
 FT CABOHD 139 139 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 346 AA; 38006 MW; 473DD2FLASBF89DE CRC64;  
 Query Match 25.4%; Score 623; DB 1; Length 346;  
 Best Local Similarity 39.2%; Pred. No. 2, 1e-34;  
 Matches 143; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 62 CPRAAPGPHSSPNPEFYLYLVGAQGRCEAPAPNLLTCDRPPDLRFETIKFOEYS 121  
 DB 64 CPRAEAGRP-----YVYLYLVPRPQAACSTVADPNVLVCNRPQGRIRITIFQES 118  
 QY 122 PNLNGEFPNSHDYIYIATSDTRBGLSLQGVCLTRGMKTLVAVGQSPGCAVPRKPV 181  
 DB 119 PNYMGLEFKKHHDYITTSNSGLBGLERBEGVCRTRMKIKMVGQDPNATPEQLTT 178  
 QY 182 SEWMERDRGAASLE-PEKENTLPGDPTSNATSRGAEGFLPPSPNPAVGAAGLA---- 236  
 DB 179 SRSKADNTVMKAAQAPSGSRGSLSDSGKHTVNOEKSQ-----GASGSSGDP 231  
 QY 237 -----LLLVGAAGA-----MWRPRRAKPSSESHPPGSGFGSLGL 277  
 DB 232 GPENSKVAFPAVAGAGCVIFLLIIFLVTLTKLKRHRKHTQ-----RAAALSL 282  
 QY 278 ---GGGGMGGRBAPGELGIALRGGAADPPCPHYEKSGDYGHVYIVQDGPPOSF 333  
 DB 283 STLASPKGSGGATGTEPDDITPLR---TTNNVCPHYEKSGDYGHVYIVQENPPOSF 339  
 QY 334 PNYY 338  
 DB 340 ANIYY 344  
 RESULT 5  
 EFB2 HUMAN STANDARD; PRT; 333 AA.  
 ID EFB2 HUMAN STANDARD; PRT; 333 AA.  
 AC P52759;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ephrin-B2 precursor (Eph-related receptor tyrosine kinase ligand 5)  
 DE (LEK-5) (HTK ligand) (HTK-L).  
 GN EFB2 OR EPLG5 OR LEK5 OR HTKL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eupheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid9606;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=96145238; PubMed=8559144;  
 RA Cerretti D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,  
 RA Maraskovsky E., Park L.S., Lyman S.D., Copeland N.G., Gilbert D.J.,  
 RA Jenkins N.A., Fletcher R.A.;  
 RT Isolation of LEK-5: a ligand of the eph-related receptor tyrosine  
 RT kinases";  
 RL Mol. Immunol. 32:1197-1205(1995).  
 RN [2]  
 RA SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RX MEDLINE=9519254; PubMed=7534404;  
 RA Bennett B.D., Zeigler F.C., Gu Q., Rendly B., Goddard A.D.,  
 RA Gillett N., Matthews W.;  
 RT "Molecular cloning of a ligand for the Eph-related receptor protein-  
 RT tyrosine kinase Hck";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).  
 RN [3]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=98192220; PubMed=9533549;  
 RA Vogt T., Stolz W., Welsh J., Jung B., Kerbel R.S., Kobayashi H.,  
 RA Landthaler M., McEllelland M.;  
 RT "Overexpression of Lerk-5/Eplg5 messenger RNA: a novel marker for  
 RT increased tumorigenicity and metastatic potential in human malignant  
 RT melanomas";  
 RL Clin. Cancer Res. 4:791-797(1998).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF  
 CC LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB4 AND EPHA3.  
 CC -1- TISSUE SPECIFICITY: LUNG AND KIDNEY.





Db 327 PANIY 332

RESULT 7

EFB2 BRARE STANDARD; PRT; 332 AA.

ID EFB2 BRARE 073874;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Ephrin-B2 precursor.

GN EFB2 OR EFB2A.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;

OC Cyprinidae; Danio.

OC NCBI\_TaxID=7955;

OK NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96438455; PubMed=9765210;

RA Durbin L., Brennan C., Shiomi K., Cooke J., Barrios A.,

RA Shammugalingam S., Guthrie B., Lindberg R., Holder N.;

RT "Eph signaling is required for segmentation and differentiation of the somites."

RL Genes Dev. 12:3096-3109(1998).

CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PMW: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE CYTOPLASMIC DOMAIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AJ004863; CAA06166.1; -

DR ZFIN; ZDB-GENE-980415-67; efmb2a.

DR InterPro; IPR001799; Ephrin.

DR Pfam; PF00812; Ephrin; 1.

DR PRINTS; PR01347; Ephrin.

DR PRODOM; PD002533; Ephrin; 1.

DR PROSITE; PS01299; EPHRIN; 1.

KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;

KW Signal; Phosphorylation.

FT SIGNAL 1 24

FT CHAIN 1 25 332

FT DOMAIN 25 225

FT TRANSEM 226 246

FT DOMAIN 247 332

FT DOMAIN 330 332

SO SEQUENCE 332 AA; 36724 MW; 189ED82372C71C8B CRC64;

Query Match 25.2%; Score 617.5; DB 1; Length 332;

Best Local Similarity 41.8%; Pred. No. 4.6e-34;

Matches 143; Conservative 54; Mismatches 106; Indels 39; Gaps 12;

Db 14 GALLLGLVGLVSGLSLEPVYVNSANKRFQAEGLVYLPQIGDLDDLCPARAPGPHSS 73

Db 11 GVLVACKNSRALIDSLITWNTNTYFVPGQGLVLPQIGDKMDIVCPVVE--GGSM 67

Qy 74 PVVEFYKYLIVGAGRCCEAPAPVNLITCDPDLDFRIFKEQSEYSPNMGHEFRSH 133

Db 68 BEVEYKYLIVPPLBGLKSCQYTKADTPLNLCVQKPDQVFKTLKQBSFNMGLEFFRGK 127

Qy 134 DYVIAITSDGREGIESAQGVCLTRGKVLTRVQSGRGGAIVRRK-PVSEMPNERDGA 192

Db 128 DYIITISNGTWEGIDNGEGVCKTKSKIMKVGQNSPDISPKDYFTSPPGHPLDG 187

Qy 193 AHS-----LEP-----GRNLPDPTSNATSRGAGSLPSPSMVAVAGAGLALLLGLV 242

Db 188 KDSKNEVLKPDASPHGECK--GDGNKSSSVIGSEVAL-----FACIASASVITIIIIIML 241

Qy 243 AGAGAMCWRRRRAKPSSESRHPGSGFG-----RCGSLGCGGGGAGREAPGELGIA 296

Db 242 VFL--LTKYRRRRRGRGS-PQHAATTLSTLSTLTPKRGDS-----GNNNG--SRPSITIF 291

Qy 297 LRGGGAADPPFCYHYKVSQDYGHVYIVQDGPQSPNNIY 338

Db 292 LR--TADSVFCYHYKVSQDYGHVYIVQEMPQSPANNIY 330

RESULT 8

EFB1 MOUSE STANDARD; PRT; 345 AA.

ID EFB1 MOUSE

AC P52795;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Ephrin-B1 precursor (Eph-related receptor tyrosine kinase ligand 2)

DE (Ephr-2) (Eph ligand) (Eph-L) (STRAL protein) (CEK5 receptor ligand)

DE (CEK5-L).

GN EFB1 OR EPLG2 OR LERK2 OR STRAL OR EPL2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=129/Sv;

RX MEDLINE=95203867; PubMed=7896266;

RA Fletcher F.A., Renshaw B., Hollingsworth T., Baum P., Lyman S.D.,

RA Jenkins N.A., Gilbert D.J., Copeland N.G., Davison B.L.;

RT "Genomic organization and chromosomal localization of mouse Eplg2, a gene encoding a binding protein for the receptor tyrosine kinase elk."

RL Genomics 24:127-132(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=95377533; PubMed=7649373;

RA Boulier P., Ouled-Abdelghani M., Vicaire S., Garnier J.M.,

RA Schudaur B., Dole P., Chambon P.;

RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in P19 embryonal carcinoma cells and characterization of a novel mouse gene, Stral (mouse LERK-2/Eplg2)."

RL Dev. Biol. 170:420-433(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=95014510; PubMed=7929389;

RA Shao H., Lou L., Pandey A., Pasquale E.B., Dixit V.M.;

RT "cDNA cloning and characterization of a ligand for the CEK5 receptor protein-tyrosine kinase."

RL J. Biol. Chem. 269:26606-26609(1994).

RN [4]

RP FUNCTION.

RX MEDLINE=20171264; PubMed=10704386;

RT "Imodi R., Wideman C., Kaprielian Z.;

RT "Complementary expression of transmembrane ephrins and their receptors in the mouse spinal cord: a possible role in constraining the orientation of longitudinally projecting axons."

RL Development 127:1397-1410(2000).

CC -1- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF LONGITUDINALLY PROJECTING AXONS.

CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB1 AND EPHB1.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS, SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE PERIOD OF COMMISSURAL AXON PATHFINDING.

```

CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC CYTOPLASMIC DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07602; AAC53247.1; -
DR EMBL; U07598; AAC53247.1; JOINED.
DR EMBL; U07599; AAC53247.1; JOINED.
DR EMBL; U07600; AAC53247.1; JOINED.
DR EMBL; U07601; AAC53247.1; -
DR MGD; MGI:102708; Efnb1.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin.1.
DR PRINTS; PR01347; EPHRIN.
DR PRODOM; PD002533; Ephrin.1.
DR PROSITE; PS01299; EPHRIN.1.
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW Signal; Phosphorylation.
FT CHAIN 1 24
FT DOMAIN 25 345
FT TRANSMEM 237 257
FT DOMAIN 258 345
FT DOMAIN 343 345
FT CARBOHYD 139 139
FT CONFLICT 90 90
SQ SEQUENCE 345 AA; 37859 MW; 8C96FD3DC5C6C405 CRC64;

Query Match 24.7%; Score 604.5; DB 1; Length 345;
Best Local Similarity 37.9%; Pred. No. 3.5e-33;
Matches 136; Conservative 51; Mismatches 107; Indels 65; Gaps 10;

QY 15 ALLLLGLVGLVSG--SLEPVYVNSANKRFQAGGYLYVQIGDRLLDLCPPRAPPGRHS 72
DB 15 AMVVLTLCLRLATPLANKLEPVSWSLNPKFLSGKGLVTPKIGDKLDITCPRAABGRP-- 72
QY 73 SPNVEFYKLYVGAQGRCEAPRPNLLTCDRPLDRLFTIKFOEYSPNLMGHEFRSH 132
DB 73 ---YEVYKLYLVPEQAAACSTYLDENVLVTCNKPQGEIRFTIKFOEFSPTMGLEFKKY 129
QY 133 HDYIITATSDGTREGLESLOGVCLTRGKVKLLRVGQSPRGAVPRKPYSEMMEERDGA 192
DB 130 HDYIITATSDGTREGLESLOGVCLTRGKVKLLRVGQSPRGAVPRKPYSEMMEERDGA 192
QY 193 AHSLEPEKENTLPDPTSNATSRGABGP-----LPPSPMAVAGAAG-----LA 236
DB 189 VKT-----ATQAPGRSGQSDSKHETVNOQEKGSGFGAGSGSDSDSFNSK 236
QY 237 LLLLVGAVAGAG-----WCWRRRAKPSSESHRPGSGRGSGL-----CG 279
DB 237 VALFAVAGVCVFLIIITFLVTLTKLKRHRKHTQ-----RAAALSTSLASP 287
QY 280 GCGMGREAPFEGELIALRGGAADPPFCPHYEKVSGDYGVYIVODPSPSPNITY 338
DB 288 KGGSGTGAPEPSDIIIFLR---TTENNVCPHYEKVSGDYGVYIVQENRPPSPNITY 343

RESULT 9
EFBI RAT STANDARD; PRT; 345 AA.
AC P52756;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DE 30-MAY-2000 (rel. 39, Last annotation update)
DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)

```

```

DE (LERK-2) (ELK ligand) (ELK-L).
GN EPHB1 OR EPIC2 OR LERK2.
OS Rattus norvegicus (Rat).
CC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95022634; PubMed=7936648;
RA Fletcher F.A., Carpenter M., Shilling H., Baum P., Ziegler S.,
RA Lyman S.D., Beckmann M.P.,
RA "LERK-2, a binding protein for the receptor-tyrosine kinase Elk, is
RT evolutionarily conserved and expressed in a developmentally regulated
RT pattern.";
RL Oncogene 9:3241-3248(1994).
CC -1- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL
CC AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE
CC ORIENTATION OF LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB3 (PREFERRED),
CC EPHB1 AND EPHB2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07560; AA53092.1; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin.1.
DR PRINTS; PR01347; EPHRIN.
DR PRODOM; PD002533; Ephrin.1.
DR PROSITE; PS01299; EPHRIN.1.
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW Signal; Phosphorylation.
FT CHAIN 1 24
FT DOMAIN 25 345
FT TRANSMEM 237 257
FT DOMAIN 258 345
FT DOMAIN 343 345
FT CARBOHYD 139 139
SQ SEQUENCE 345 AA; 37951 MW; 1B3045C5C7358F7E CRC64;

Query Match 24.5%; Score 599.5; DB 1; Length 345;
Best Local Similarity 37.6%; Pred. No. 7.4e-33;
Matches 135; Conservative 52; Mismatches 107; Indels 65; Gaps 10;

QY 15 ALLLLGLVGLVSG--SLEPVYVNSANKRFQAGGYLYVQIGDRLLDLCPPRAPPGRHS 72
DB 15 AMVVLTLCLRLATPLANKLEPVSWSLNPKFLSGKGLVTPKIGDKLDITCPRAABGRP-- 72
QY 73 SPNVEFYKLYVGAQGRCEAPRPNLLTCDRPLDRLFTIKFOEYSPNLMGHEFRSH 132
DB 73 ---YEVYKLYLVPEQAAACSTYLDENVLVTCNKPQGEIRFTIKFOEFSPTMGLEFKKY 129
QY 133 HDYIITATSDGTREGLESLOGVCLTRGKVKLLRVGQSPRGAVPRKPYSEMMEERDGA 192
DB 130 HDYIITATSDGTREGLESLOGVCLTRGKVKLLRVGQSPRGAVPRKPYSEMMEERDGA 192
QY 193 AHSLEPEKENTLPDPTSNATSRGABGP-----LPPSPMAVAGAAG-----LA 236
DB 189 VKT-----ATQAPGRSGQSDSKHETVNOQEKGSGFGAGSGSDSDSFNSK 236
QY 237 LLLLVGAVAGAG-----WCWRRRAKPSSESHRPGSGRGSGL-----CG 279

```

```

Db 237 VALFAVAGCVIFLIIIFLVLTKLRKRKHTQQ-----RAALSLTLASP 267
Qy 260 GGMGPREAREPEGLIALRGGAADPPCPHEKXSGYGHPIYVQDGPQSPNITY 338
Db 268 KCDSTAGTAEPSDIIIFLR--TTENNYPHEKXSGYGHPIYVQEMPQSPNITY 343

RESULT 10
EFB1_XENLA STANDARD; PRT: 327 AA.
AC 01097:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)
DE (LEKR-2) (ELK ligand) (ELK-L) (XLERK).
GN EFNB1 OR EPLG2 OR LERK2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9716777; PubMed=9174051;
RA Jones T.L., Karavanova I., Chong L., Zhou R.P., Daar I.O.;
RT "Identification of Xlerk, an Eph family ligand regulated during
RT mesoderm induction and neurogenesis in Xenopus laevis.";
RT Oncogene 14:2159-2166(1997).
CC -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPING MESENCHYMAL AND
CC NERVOUS TISSUE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT LOW LEVELS IN MOST
CC ADULT TISSUES WITH INCREASED LEVELS OBSERVED IN THE KIDNEY,
CC OOCYTES, OVARY AND TESTIS.
CC -1- PUT: INDICIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U31427; AAC35995.1; -
CC InterPro: IPR001799; Ephrin.
CC Pfam: PF00812; Ephrin.1.
CC PRINTS: PR01347; EPHRIN.
CC PRODOM: PD002533; Ephrin; 1.
CC PROSITE: PS01299; EPHRIN; 1.
CC Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
CC signal; Phosphorylation.
CC FT SIGNAL 1 20 POTENTIAL.
CC FT CHAIN 21 327 EPHRIN-B1.
CC FT DOMAIN 21 225 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 226 246 POTENTIAL.
CC FT DOMAIN 247 327 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 325 327 PDZ RECOGNITION MOTIF (POTENTIAL).
CC FT CARBOHYD 131 131 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 202 202 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 327 AA; 36621 MW; 71230CE7F6BE5974 CRC64;

Query Match 24.1%; Score 591; DB 1; Length 327;
Best Local Similarity 39.5%; Pred. No. 2.6e-32;
Matches 144; Conservative 43; Mismatches 100; Indels 78; Gaps 12;

```

```

Qy 66 RPPGPHSSPNVEYKLYVGAAGRCCEAPAPNLLTCDRPDLURFTIKOEYSNIM 125
Db 63 QP-----YERYKLYMRDQLEACSTVIDNPVLTQNGKEYRFTIKOEYSNIM 114
Qy 126 GHEFRSHHDYIIATSDGTREGLESLOGGVCLTRGKVKLLRVGQSPRGAVPRKPYSEMP 185
Db 115 GLEFRNODYIITSNSTSLQGLENRGGVCTQSRMKIIMKYGQDP--NAVPEQILT--- 169
Qy 186 MERDRGAASLSPEKGNLEPDPPTSNATSGA--EGPLPP-----SPPAAGAAGLA 236
Db 170 -----TTPRSKE--ADNYGKIATGPNNGVQNGKSDTNLSDKPTGRGWYDGF 217
Qy 237 LLLLVGAAGAGAMC-----NRRRAKPSK-----SPHPPGSGFRGG 273
Db 218 NSKIAFAAIIAGAGCVIFLIIIFVLVLTKIKRKRKHTQQAALSLTLASPCKSGNNG 277
Qy 274 SLGLGGGGMGPREAREPEGLIALRGGAADPPCPHEKXSGYGHPIYVQDGPQSP 333
Db 278 S-----EPSDIIIFLR--TTENNYPHEKXSGYGHPIYVQEMPQSP 320
Qy 334 PNITY 338
Db 321 ANITY 325

RESULT 11
EFB2_BRARE STANDARD; PRT: 195 AA.
ID EFB2_BRARE
AC P79727;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE (LEKR-6) (ELF-1) (ZEPHRL3).
GN EFN2 OR EPL6 OR LERK6.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthys; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97195707; PubMed=9043080;
RA Brenman C., Monschau B., Lindberg R., Guthrie B., Drescher U.,
RA Bonhoeffer F., Holder N.;
RT "Two Eph receptor tyrosine kinase ligands control axon growth and may
RT be involved in the creation of the retinotectal map in the
RT zebrafish ";
RT Development 124:655-664(1997).
CC -1- FUNCTION: CONTROL AXON GROWTH AND MAY BE INVOLVED IN THE CREATION
CC OF THE RETINO-TECTAL MAP.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN THE EMBRYO.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE PRESUMPTIVE MIDBRAIN OF
CC EXPRESSED THROUGHOUT THE MIDBRAIN INCLUDING THE REGION OF THE
CC PRESUMPTIVE TECTUM. AT LATER STAGES IT IS EXPRESSED IN A GRADED
CC FASHION THROUGHOUT THE TECTUM.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y09668; CAA70863.1; -
CC ZFIN: ZDB-GENE-990415-66; efn2.

```

DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.1.  
 DR PRINTS: PRO1347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin.1.  
 DR PROSITE: PS01299; EPHRIN.1.  
 KW Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 195  
 FT CARBOHYD 32 32 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 195 AA; 2268 MM; 9EB284FEB61DC42 CRC64;  
 Query Match 7.6%; Score 185; DB 1; Length 195;  
 Best Local Similarity 29.9%; Pred. No. 9.3e-06;  
 Matches 63; Conservative 21; Mismatches 71; Indels 56; Gaps 7;  
 QY 33 VYNNSSANKRPGAGGYVLPQIDRLDLCPRARPPGSSPVYFYKYLYGAGAGRC 92  
 DB 29 VYNNSSNSRFW-QGEYVAVSINDYDVCYFYESPQFS-RMEYXILFMVNHGGLYTC 85  
 QY 93 EAPPAFLITLTCR---PDLRLFTIKFOEYSPNLGHSFRSHHDYIATSDGTREGLE 149  
 DB 86 EHMWRGKMEKRCRPOSDFPLAFSEKFOLEFTRPFGFEFRPCHETYYISSTHPNAGRP 145  
 QY 150 SLOGGVCLTRGKAVYLRLVGSPRGCAVPRKPYSEMPMERDRGAHSLFPGKENLPDPTS 209  
 DB 146 CLK-----LKYY-----KPTSS-----GYSEPEPFLTD 169  
 QY 210 NATSRGABEPPLPPSPAVAGAGLALLL 240  
 DB 170 QSORCAGHDFC-----LAVLML 186  
 RESULT 12  
 EF33 HUMAN STANDARD; PRT; 238 AA.  
 AC P52757;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ephrin-A3 precursor (EPH-related receptor tyrosine kinase ligand 3)  
 DE (LERK-3) (EHK1 ligand) (EHK1-L).  
 GN EPN3 OR EPLG3 OR LERK3 OR EFL-2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX MEDLINE=95140419; PubMed=7838529;  
 RX Kozlosky C.J., Maraskovsky B., McGraw J.T., Vandenbos T.,  
 RX Teepe W., Lymn S.D., Strinivasan S., Fletcher F.A., Gayle R.B. III,  
 RX Cerecetti D.P., Beckmann M.P.;  
 RA "Ligands for the receptor tyrosine kinases hek and elk: isolation of  
 RT cDNAs encoding a family of proteins";  
 RN Oncogene 10:239-306(1995).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95063919; PubMed=7973638;  
 RA Davis S., Gale N.W., Aldrich T.H., Malsompierre P.C., Lhotak V.,  
 RA Pawson T., Goldfarb M., Yancopoulos G.D.;  
 RA "Ligands for EPH-related receptor tyrosine kinases that require  
 RT membrane attachment or clustering for activity";  
 RL Science 266:816-819(1994).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, SKELETAL MUSCLE, SPLEEN,  
 CC THYMUS, PROSTATE, TESTIS, OVARY, SMALL INTESTINE, AND PERIPHERAL  
 CC BLOOD LEUKOCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U14187; AAC50078.1; -;  
 DR EMBL: L37360; AAA52368.1; -;  
 DR Genbank: HGNC:3223; EPN3.  
 DR MIM: 601381; -;  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.1.  
 DR PRINTS: PRO1347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin.1.  
 DR PROSITE: PS01299; EPHRIN.1.  
 KW Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 238  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 67 67 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 100 100 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CONFLICT 71 74 MISSING (IN REF. 2).  
 SQ SEQUENCE 238 AA; 26350 MM; 8FPD6ABEF33PDA CRC64;  
 Query Match 7.3%; Score 179; DB 1; Length 238;  
 Best Local Similarity 28.4%; Pred. No. 2.9e-05;  
 Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;  
 QY 7 GRGVRGALLLGLVGLSLGSLPYNNSSANKRPGAGGYVLPQIDRLDLCR-R 64  
 DB 24 GFG-----GALG-----NRAVYNNSSNOHLRR-GYVQVNVNDYLIYCPHYN 67  
 QY 65 ARPPGSSP---NYEYKLYLYGAGRCRCEAPPAFLITLTCRBDL---DLRFTIKF 117  
 DB 68 SSGVGPGAGPGPGGAGQYLYWVSRYGRTCNASQGF-RMECNRPAPHSPIKFEKF 126  
 QY 118 QEYSPNLGHSFRSHHDYIATSDGTREGLESLGQGVCLTRGKAVYLRLVGSPRGCAV 177  
 DB 127 QRYAFSLGIEFHAGHEYYIISTPTNHL-----WKCLR-WKVFVCCASTSHSG--- 174  
 QY 178 RKPVSSEMP-----MERDRGAHSLF-----PGKENLP 204  
 DB 175 EKVPTPLPQFTMGPNVKNVLEDPESNPQVPLKESISGTSKRRLHP 223  
 RESULT 13  
 EF32 MOUSE STANDARD; PRT; 209 AA.  
 AC P52801;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)  
 DE (LERK-6) (ELF-1) (CEK7-ligand) (CEK7-L).  
 GN EPN2 OR EPLG6 OR LERK6 OR ELF1 OR EPL6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX MEDLINE=95181289; PubMed=7876076;  
 RX STRAIN=Swiss Webster; TISSUE=Brain;  
 RX MEDLINE=95007776; PubMed=7522971;  
 RA Cheng H.J., Flanagan J.G.;  
 RA "Identification and cloning of ELF-1, a developmentally expressed  
 RT ligand for the Mek4 and Sek receptor tyrosine kinases";  
 RL Cell 79:157-166(1994).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95181289; PubMed=7876076;  
 RA Shao H., Lou L., Pandey A., Verderame M.F., Steyer D.A., Dixit V.M.;  
 RA "CDNA cloning and characterization of a Cdk7 receptor  
 RT protein-tyrosine kinase ligand that is identical to the ligand  
 RT (ELF-1) for the Mek-4 and Sek receptor protein-tyrosine kinases";

```

RL J. Biol. Chem. 270:3467-3470(1995).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC EPHA5.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14841; AAA53636.1; -
DR EMBL; U14752; AAA68520.1; -
DR MGI; MGI:102707; Etna2.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; EPHRIN.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
DR Glycoprotein; GPI-anchor; Signal.
KM SIGNAL
FT CHAIN 1 20
FT CARBOHYD 21 209 EPHRIN-A2.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 209 AA; 23586 MW; F1997545F25B9ABC CRC64;

Query Match 7.2%; Score 176; DB 1; Length 209;
Best Local Similarity 29.3%; Pred. No. 3.9e-05;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

OY 33 VYVNSANKRFQAE-----GGVLYVQIGRDLDLCPRARPPGPHSSPNVYEFKLYLVGCA 87
DB 35 VYVNSNRPFOVSAGDGGYVVEVSINDYDIPCYPHGAPLP-PAERMEHYIIVYNGE 93
OY 88 QGRRCCEAPAPNLLITCDRPDI--DLRFTIKFQESPNLWGHFRSHHDYIATSDGT 144
DB 94 GHASCDHROGRGFKWECNRPAAPGGLFKSEKFLQFTPFSLGFERRPDHEYYIATP-- 151
OY 145 REGLESLOGGCVLTIRGMKYLRLVGVGSPRGAVPRKPVSEPMEDRGAHSLFEGKENLP 204
DB 152 ---DNLVDRPCLR--LKYVVR-----PTNETLY 174
OY 205 GDP---TSNATSRGAG 218
DB 175 EAPBPIPTSSSCSGLG 192

RESULT 14
EPA2_HUMAN STANDARD; PRT; 213 AA.
ID EPA2_HUMAN STANAD; PRT; 213 AA.
AC O43921; 076020;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE (Ephr-6) (HEK7-ligand) (HEK7-L).
GN EFNA2 OR EPLG6 OR LERK6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=98126446; PubMed=9465306;
RA Cerretti D.P., Nelson N.;
RT "Characterization of the genes for mouse LERK-3/ephrin-A3 (Ep13),
RT mouse LERK-4/ephrin-A4 (Ep14), and human LERK-6/ephrin-A2 (EPLG6);
RT conservation of intron/exon structure."

```

```

RL Genomics 47:131-135(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Garcia J., Dangnan L., Poundstone P.,
RA Christensen M., Georges A., Ayala J., Liu S., Attix C., Andreise T.,
RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Mongomery M., Ow D., Nolan M., Tyong S., Kobayashi A., Olsen A.O.,
RA Carraro A.V.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99045414; PubMed=9826538;
RA Aasheim H.C., Pedutoeur P., Grosgeorge J., Logtenberg T.;
RT "Cloning, chromosomal mapping, and tissue expression of the gene
RT encoding the human Eph-family kinase ligand ephrin-A2."
RL Biochem. Biophys. Res. Commun. 252:378-382(1998).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC EPHA5.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U92896; AAC39577.1; -
DR EMBL; U92893; AAC39577.1; JOINED.
DR EMBL; U92894; AAC39577.1; JOINED.
DR EMBL; AC004258; AAC04896.1; -
DR EMBL; AJ007292; CAA07435.1; -
DR Genew; HGNC:3222; EFNA2.
DR MIM; 602756; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; EPHRIN.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
KM Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 24
FT CHAIN 25 213 EPHRIN-A2.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 6 6 R -> A (IN REF. 3).
FT CONFLICT 25 26 RA -> PP (IN REF. 3).
FT CONFLICT 29 30 AA -> RR (IN REF. 3).
SQ SEQUENCE 213 AA; 23878 MW; 33C9PFA168B2D0 CRC64;

Query Match 7.2%; Score 175.5; DB 1; Length 213;
Best Local Similarity 36.8%; Pred. No. 4.3e-05;
Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;

OY 33 VYVNSANKRFQAE-----EGVLYVQIGRDLDLCPRARPPGPHSSPNVYEFKLYLVGCA 87
DB 39 VYVNSNRPFOVSAGDGGYVVEVSINDYDIPCYPHGAPLP-PAERMEHYIIVYNGE 97
OY 88 QGRRCCEAPAPNLLITCDRPDI--DLRFTIKFQESPNLWGHFRSHHDYIATSDGT 141
DB 98 GHASCDHROGRGFKWECNRPAAPGGLFKSEKFLQFTPFSLGFERRPDHEYYIATP 154

RESULT 15
EPA2_CHICK STANDARD; PRT; 200 AA.
ID EPA2_CHICK STANAD; PRT; 200 AA.

```



```

AC P52802;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
GN (LERK-6) (ELF-1).
DE EFNA2 OR EPLG6 OR LERK6 OR ELF1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95360981; Pubmed=7634327;
RA Cheng H.J., Nakamoto M., Bergmann A.D., Flanagan J.G.;
RT "Complementary gradients in expression and binding of ELF-1 and Mek4
in development of the topographic retinorectal projection map."
RL Cell 82:371-381(1995).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EFHA3, EFHA4 AND
CC EPHAS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A GRADIENT ACROSS THE TECTUM
CC BEING MORE STRONGLY EXPRESSED AT THE POSTERIOR POLE.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: LA0932; AAC42229.1; -
DR InterPro: IPR001799; Ephrin.
DR Pfam: PF00812; Ephrin.1.
DR PRINTS: PR01347; EPHRIN.
DR ProDom: PD002533; Ephrin.1.
DR PROSITE: PS01299; EPHRIN.1.
KW Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 22
FT CHAIN 1 220
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 200 AA; 23049 MW; 8FAB1A5E45ED96 CRC64;

Query Match 7.0%; Score 172; DB 1; Length 200;
Best Local Similarity 35.3%; Pred. No. 6.9e-05;
Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;

QY 15 ALLLGLVGLVSGLSLEP-----VYNSANKRFOAGGYLYPQIGRLDILCPRA 65
DB 7 ALLLAIVAVCVWSDDPKGVISDRAYAVYMNSTNPFH-RGVYTVESINDVLDICPHY 64
QY 66 RPPGSSSPVYEFYKLYVGAGRCCEAPPAVLLTCRDPDL---DLRFITKQETSF 122
DB 65 EEPFLP--AERMEYVLYMNYEGHACDHRQKQFKRMECNRPDSGPKLFESEKQLFPT 122
QY 123 NLWGHEFRSHDYIATIS 141
DB 123 FSLGFEFRGHEHYIISAS 141

RESULT 16
EFNA5 CHICK STANDARD; PRT; 228 AA.
ID EFNA5 CHICK
AC P52804;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

```

DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
GN (LERK-7) RAGS protein.
DE EFNA5 OR RAGS.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95360980; Pubmed=7634326;
RA Drescher U., Kremsner C., Handwerker C., Loschinger J., Noda M.,
RA Bonhoeffer F.;
RT "In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa
retinal protein related to ligands for Eph receptor tyrosine
kinases."
RL Cell 82:359-370(1995).
CC -1- FUNCTION: INDUCES GROWTH CONE COLLAPSE AND REPULSION OF RETINAL
CC GANGLION CELL AXONS.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A GRADED FASHION ACROSS THE
CC TECTUM BEING MORE STRONGLY EXPRESSED TOWARDS THE POSTERIOR POLE.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X90377; CA63027.1; -
DR InterPro: IPR001799; Ephrin.
DR Pfam: PF00812; Ephrin.1.
DR PRINTS: PR01347; EPHRIN.
DR ProDom: PD002533; Ephrin.1.
DR PROSITE: PS01299; EPHRIN.1.
KW Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 20
FT CHAIN 1 228
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 228 AA; 26206 MW; 56D8E4FEDBCF18AD CRC64;

Query Match 7.0%; Score 170.5; DB 1; Length 228;
Best Local Similarity 28.9%; Pred. No. 9.9e-05;
Matches 73; Conservative 30; Mismatches 91; Indels 59; Gaps 13;

QY 16 LLLIGVGLVSGLSLEP-----VYNSANKRFOAGGYLYPQIGRLDILCPRA 64
DB 6 MLLLAVALMWCVAGDEPRKAVADRYAVYMNSTNPFQ-QGVYHIDVCINDVLDVCPH 64
QY 65 ARPSPSSSPVYEFYKLYVG-----GAQGRCEAPPAVLLTCRDPDLR 112
DB 65 YEDVDEPKT--ERYVLYMNYNFDGYSCHDISKGFKMECNRHSFV-----GPKL 113
QY 113 FTIRQESPNLWGHEFRSHDYIATIS---DTEGLESLLGGVCLTGKMYLLRVGQ 169
DB 114 FSEKFLFTFFSLGFEFRPREYFYISSAIPDNGRS-----CLK--LKVFAV--- 159
QY 170 SPRGAVPRKPYSEMWERGAAHSLPEKELPDDPTSNATSRGAEGLLPPSPM--A 227
DB 160 -PANSCKTIGVHDVRFVDVNDKVENSLDEPADTV---RESAPSRG-ENNAQTPIRIL 214
QY 228 VAGAAGLALLL 240
DB 215 LATLLFLATLLI 227

RESULT 17
EFNA4 HUMAN

```



```

ID EF4_HUMAN STANDARD; PRT; 201 AA.
AC P52798; 095457;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ephrin-A4 precursor (EPH-related receptor tyrosine kinase ligand 4)
DE (LERK-4).
GN EFNA4 OR EPLG4 OR LERK4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=95140419; PubMed=7838529;
RX Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,
RX Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,
RX Cerretti D.P., Beckmann M.P.;
RA "Ligands for the receptor tyrosine kinases hek and elk: isolation of
RT cDNAs encoding a family of proteins."
RL Oncogene 10:299-306(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=B-cell;
RX MEDLINE=20076261; PubMed=10607706;
RA Aasheim H.C., Munthe E., Punderud S., Smeland E.B., Beiske K.,
RA Logtenberg T.;
RT "A splice variant of human ephrin-A4 encodes a soluble molecule that
RT is secreted by activated human B lymphocytes."
RL Blood 95:221-230(2000).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INTERACTION BETWEEN ACTIVATED B
CC LYMPHOCYTES AND DENDRITIC CELLS IN TONSILS.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC OR SECRETED (DEPENDING ON THE ISOFORM).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/GPI-anchored (shown here) and
CC 2/secreted; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE ADULT SPLEEN, LYMPH NODE,
CC PROSTATE, OVARY, SMALL INTESTINE, AND COLON, AND IN FETAL HEART,
CC LUNG, LIVER AND KIDNEY. ALSO DETECTED IN HEMATOPOIETIC CELL LINES.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14188; AAC50079.1; -
DR EMBL; AJ006352; CA06992.1; -
DR EMBL; AJ006353; CA06993.1; -
DR Gene; HGNC:3224; EFNA4.
DR MIM; 601390;
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; EPHRIN.
DR PRODOM; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
KW Glycoprotein; GPI-anchor; Signal; Alternative splicing.
FT SIGNAL 1 22
FT CHAIN 23 170
FT PROPEP 171 201
FT CARBOHYD 33 33
FT LIPID 170 170
FT VARSPIC 157 201
FT SEQUENCE 201 AA; 22386 MW; A3B8D5443A9AF28D CRC64;
Query Match 6.9%; Score 169.5; DB 1; Length 201;
Best Local Similarity 29.9%; Pred. No. 0.0001;
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

```

```

CY 24 LVGLSLI-EPVYNSANKRFOAGGYLYVPOIGDRDLLCPRAPPPGSSPNYEFYKY 82
DB 20 LRGSSLRHVYVWNSNPRLLRGDAVVEGLNDYLDIVCPHYEGGPEGP--ETFALY 76
CY 83 LVGAGQRREAR-PAPNLLITCDRDDLRLFRFKQESPNLMGHEFSSHDDYIALS 141
DB 77 VMWMPGYESCQAGPAPYAKRWVCSLPFGHVQSEKIQRTFSLGFEFLPGETTYISVP 136
CY 142 DGTREGLIESLQGCVCITRGKVLRLRVGQSPRGAVPRKPVSEMPERDRGAASLEPGKE 201
DB 137 --TPE-----SSGQL-----RLGVSCCKRKESAHV----- 164
CY 202 NLPGDPNSNATN--RQAGPLPPSPVPAVGAAGCALILL 240
DB 165 ---GSPSGTSGMRGDTFSP-----LCILL 189
RESULT 18
EFAS_BRARE STANDARD; PRT; 228 AA.
ID EFAS_BRARE
AC P79728;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE (LERK-7) (AL-1) (ZEPH14).
GN EFNA5 OR EFNA5B OR AL1 OR EPLG7 OR LERK7.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID:7955;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Embryo;
RX MEDLINE=97195707; PubMed=9043080;
RX Brennan C., Munschau B., Lindberg R., Guthrie B., Drecher U.,
RX Bonhoeffer F., Holder N.;
RT "Two Eph receptor tyrosine kinase ligands control axon growth and may
RT be involved in the creation of the retinotectal map in the
RT zebrafish."
RL Development 124:655-664(1997).
CC -1- FUNCTION: CONTROL AXON GROWTH AND MAY BE INVOLVED IN THE CREATION
CC OF THE RETINO-TECTAL MAP.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -1- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN THE EMBRYO.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE PRESUMPTIVE MIDBRAIN OF
CC DEVELOPING EMBRYOS FROM THE SIX-SOMITE STAGE. BY 24 HOURS IT IS
CC STRONGLY EXPRESSED IN THE MIDBRAIN CAUDAL TO THE PRESUMPTIVE
CC TECTUM. AT LATER STAGES IT IS MAINTAINED AT THE POSTERIOR MARGIN
CC OF THE TECTUM.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09669; CA170864.1; -
DR ZFIN; ZDB-GENE-980526-186; efna5b.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; EPHRIN.
DR PRODOM; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
KW Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 20
FT CHAIN 21 228
FT EPHRIN-A5.

```



```

CC THE FYN TYROSINE KINASE.
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND
CC EPHB1.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR. IT
CC IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE
CC MICRODOMAINS.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U06403; AAB60377.1; -.
DR Genbank: HGNC:3225; EFNA5.
DR MIM: 601535; -.
DR InterPro: IPR001799; Ephrin.
DR Pfam: PF00812; Ephrin.1.
DR PRINTS: PR01347; EPHRIN.
DR ProDom: PD002533; Ephrin.1.
DR PROSITE: PS01299; EPHRIN.1.
DR Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal;
KM Polymorphism.
KM FT SIGNAL 1 20
FT CHAIN 21 228
FT CARBOHYD 37 37 N-LINKED (GLCNAC...;) (POTENTIAL).
FT VARIANT 55 55 N->K (IN DBSNP:469062).
FT /FTID-VAR 012035.
SQ SEQUENCE 228 AA; 26297 MW; 6893B1C6CF357 CRC64;

Query Match 6.8%; Score 167.5; DB 1; Length 228;
Best Local Similarity 28.8%; Pred. No. 0.00016;
Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;

QY 33 VYNSANKRQAGGAYLYPQIGDRDLCPRARPPGPHSPVPEYKYLYG----- 85
DB 34 VYNSNSPRFQ-RGDYHIDVINDYDVFCPHYEDSVPEDKT--ERYVLYMVPDGSAC 90
QY 86 -----GAQGRCEAPAPANNLLTCRPPDLRFTIKFOEYSPNMGHEFSHDDYLIAT 140
DB 91 DHTSKGFKMECNRPSPN-----GPKTSEKQFLTPSLGFEFRPGREYFISS 141
QY 141 S---DGTREGLSLOGGVCITRGKYLRLVGQSPRGAVPRKYSMPMERDGAASLE 197
DB 142 AIPDNGRRS-----CLK-LKVFVR---FTNSCKMTIGVHDVDFVNDKVENSLR 186
QY 198 PGKENLPGDPTSNATSRGAGPLPPSPMPAVAGAGLALLLIGVA 243
DB 187 PADDTV---HESAPSRG-ENAAQTPIPSRL-----LAILFLLA 223

RESULT 21
EFNA5 MOUSE STANDARD; PRT; 228 AA.
AC 008543; 008544:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE (LEHR-7) (AL-1).
GN EFNA5 OR EPLG7 OR LERK7 OR EPL7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=97060319; PubMed=8903354;
RA Fleniken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;
RT "Distinct and overlapping expression patterns of ligands for

```

```

RT Eph-related receptor tyrosine kinases during mouse embryogenesis.";
RL Dev. Biol. 179:382-401(1996).
CC -1- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON FASCICULATION.
CC INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE
CC MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS
CC COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF
CC THE FYN TYROSINE KINASE (BY SIMILARITY).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND
CC EPHB1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR. IT
CC IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE
CC MICRODOMAINS (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U06654; AAB50239.1; -.
DR Genbank: U90665; AAB50240.1; -.
DR MIM: 601535; -.
DR InterPro: IPR001799; Ephrin.
DR Pfam: PF00812; Ephrin.1.
DR PRINTS: PR01347; EPHRIN.
DR ProDom: PD002533; Ephrin.1.
DR PROSITE: PS01299; EPHRIN.1.
DR Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal;
KM Alternative splicing.
KM FT SIGNAL 1 20
FT CHAIN 21 228
FT CARBOHYD 37 37 N-LINKED (GLCNAC...;) (POTENTIAL).
FT VARIANT 163 189 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 228 AA; 26339 MW; 85439F537420022 CRC64;

Query Match 6.8%; Score 167.5; DB 1; Length 228;
Best Local Similarity 28.8%; Pred. No. 0.00016;
Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;

QY 33 VYNSANKRQAGGAYLYPQIGDRDLCPRARPPGPHSPVPEYKYLYG----- 85
DB 34 VYNSNSPRFQ-RGDYHIDVINDYDVFCPHYEDSVPEDKT--ERYVLYMVPDGSAC 90
QY 86 -----GAQGRCEAPAPANNLLTCRPPDLRFTIKFOEYSPNMGHEFSHDDYLIAT 140
DB 91 DHTSKGFKMECNRPSPN-----GPKTSEKQFLTPSLGFEFRPGREYFISS 141
QY 141 S---DGTREGLSLOGGVCITRGKYLRLVGQSPRGAVPRKYSMPMERDGAASLE 197
DB 142 AIPDNGRRS-----CLK-LKVFVR---FTNSCKMTIGVHDVDFVNDKVENSLR 186
QY 198 PGKENLPGDPTSNATSRGAGPLPPSPMPAVAGAGLALLLIGVA 243
DB 187 PADDTV---HESAPSRG-ENAAQTPIPSRL-----LAILFLLA 223

RESULT 22
EFNA5 RAT STANDARD; PRT; 228 AA.
AC P97605;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)
DE (LEHR-7) (AL-1).
GN EFNA5 OR EPLG7 OR LERK7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=95267434; PubMed=7748564;  
 RA Winslow J.W., Moran P., Valverde J., Shih A., Yuan J.Q., Wong S.C.,  
 RA Tsai S.P., Gaddard A., Henzel W.J., Heitel P., Beck K.D., Carsi I.W.,  
 RT "Cloning of AL-1, a ligand for an Eph-related tyrosine kinase  
 RT receptor involved in axon bundle formation.";  
 RL Neuron 14:973-981(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Li Y.Y., McIverman C.F., Feldman A.M.,  
 RT "Etkrty, rat ligand for Eph-related receptor tyrosine kinase.";  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON FASCICULATION.  
 CC INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE  
 CC MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS  
 CC COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF  
 CC THE FYN TYROSINE KINASE (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND  
 CC EPHA1 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR. IT  
 CC IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE  
 CC MICRODOMAINS (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA AND LUNG.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U69279; AAC05801.1; -  
 DR InterPro: IPR001799; Ephrin.1.  
 DR Pfam: PF00812; Ephrin.1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin.1.  
 DR PROSITE: PS01299; EPHRIN.1.  
 KW Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 1 228 EPHRIN-A5.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 228 AA; 26358 MW; 855985532D580022 CRC64;  
 Query Match 6.8%; Score 167.5; DB 1; Length 228;  
 Best Local Similarity 28.8%; Pred. No. 0.00016;  
 Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;  
 QY 33 VYNSANKRFOAGGYVLPQIGRLDLCPRAPPGPHSSPNYFYKYLVG----- 85  
 DB 34 VYNSNSNPRFQ-RSDYHIDVCINDYLVFCPHYEDSVPEKXT-ERYVLVNVNFDGYSAC 90  
 QY 86 -----GAGRCCEAPAPNLLTCDRDLRPTIKFOEYSPNMGHEFSHDDYIAT 140  
 DB 91 DHTSKGFPRMCCNRPNSN-----GPKSEKQLTTPSLGEEFPGMEIFITSS 141  
 QY 141 S---DGTBEGLESLQGVCLTRGMKVLIRVQSPRGAVPRKPYSEMPMERDRGAHSLR 197  
 DB 142 AIPDNGRS-----CLK--LKVFVR-----PTNSCMKTIIVRDRVFNKVENSLR 186  
 QY 198 PGKENTLPDPTSNATSGABEPLPPSPMRAVAGAGLALLLLGVA 243  
 DB 187 PADTV--HESAPSRG-ENVAAGTPRIPSL-----LAILLFLLA 223

ID EFAL\_HUMAN STANDARD; PRT; 205 AA.  
 AC P20827;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
 DE (LRRK-1) (Immediate early response protein B61) (Tumor necrosis  
 DE factor, alpha-induced protein 4).  
 GN EFNA1 OR EFG1 OR LERK1 OR TNFAIP4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91042512; PubMed=2233719;  
 RA Holzman L.B., Marks R.M., Dixit V.M.,  
 RT "A novel immediate-early response gene of endothelium is induced by  
 RT cytokines and encodes a secreted protein.";  
 RL Mol. Cell. Biol. 10:5830-5838(1990).  
 RN [2]  
 RP GPI-ANCHOR.  
 RX MEDLINE=95140419; PubMed=7838529;  
 RA Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,  
 RA Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,  
 RA Cerretti D.P., Beckmann M.P.,  
 RT "Ligands for the receptor tyrosine kinases hex and elk: isolation of  
 RT cDNAs encoding a family of proteins.";  
 RL Oncogene 10:299-306(1995).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,  
 CC EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EFNA1.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- INDUCTION: BY TNF-ALPHA AND INTERLEUKIN-1 BETA.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M57730; AAA58388.1; -  
 DR PIR: A36377; A36377.  
 DR Gene: HGNC:3221; EFNA1.  
 DR MIM: 191164;  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR PRODOM: PD002533; Ephrin.1.  
 DR PROSITE: PS01299; EPHRIN.1.  
 KW Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 1 205 EPHRIN-A1.  
 FT CARBOHYD 26 26 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 205 AA; 23771 MW; 4FEFC6B4C1251A9 CRC64;  
 Query Match 6.8%; Score 166; DB 1; Length 205;  
 Best Local Similarity 27.5%; Pred. No. 0.00018;  
 Matches 46; Conservative 31; Mismatches 74; Indels 16; Gaps 5;  
 QY 18 ILGLVGLVSGLSLEPVYNSANKRFOAGGYVLPQIGRLDLCPRAPPGPHSSPN- 75  
 DB 8 ILGLCSIAADRHVFNWNSNPKFRNE-DYTIHVQLNDYDIIIPHYE--DHSVADA 63  
 QY 76 VEFYCLYLVGAGRCAPAPNLLTCDRDL---DLRTIKFOEYSPNMGHEFSH 132  
 DB 64 MEQYLLYVHEEYQLCPQSKDQYRWQCNRPSSAKHGPKSEKFRQRTPTLTKGFEKG 123  
 QY 133 HDYIATSDGTREGLESLQGVCLTRGMKVLIRVQSPRGAVPRK 179  
 DB 124 HSYIYSKPIHQHEDR-----CLRLKVTYVSGKITHSRQAHVNPQE 163

RESULT 24

ID	EPH1_RAT	STANDARD	PRT	205 AA
AC	P97553			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)			
DE	(LEKX-1) (Immediate early response protein B61)			
GN	EPHNA1 OR EPHL1 OR LERK1			
OS	Rattus norvegicus (Rat)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=Miscar; TISSUE=Brain;			
RX	MEDLINE=95405853; PubMed=7675446;			
RA	Takahashi H., Ikeda T.;			
RT	"Molecular cloning and expression of rat and mouse B61 gene;			
RT	implications on oncogenesis.";			
RL	Oncogene 11:879-883(1995).			
CC	-1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,			
CC	EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1			
CC	(BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By			
CC	similarity).			
CC	-1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D38056; BAA07242.1; -			
DR	InterPro: IPR001799; Ephrin.			
DR	Pfam: PF00812; Ephrin; 1.			
DR	PRINTS; PR01347; EPHRIN.			
DR	ProDom; PD002533; Ephrin; 1.			
DR	PROSITE; PS01299; EPHRIN; 1.			
KW	Glycoprotein; GPI-anchor; Signal.			
FT	SIGNAL 1 17			
FT	CHAIN 18 205			
FT	CARBOHYD 26 205			
FT	SEQUENCE 205 AA; 23718 MW; C86DAB3D56A6EAD CRC64;			
CC	N-LINKED (GLCNAC... ) (POTENTIAL).			
CC	-----			
CC	Query Match 6.6%; Score 162.5; DB 1; Length 205;			
CC	Best Local Similarity 25.7%; Pred. No. 0.0003;			
CC	Matches 49; Conservative 33; Mismatches 66; Indels 23; Gaps 6;			
CC	-----			
QY	18 LIGLVGLVSGSLPEPYMNSAKRFOAEGGYLYFOIGRDLDPKRRAPGPHSSPYE 77			
DB	8 LIGLCCSLAIVRHRIVFMWNSNPKFRRE-DYTVHQLNDYDILICPHYEDDSV-ADAAME 65			
QY	78 FYKLYLVGAQGRCEAPAPNILLTCRPLD---DLFTIFQCEYSENLWGHEFRSHD 134			
DB	66 RYSLVYVHEQEVYTCBPQSKQVAKNCPKSAKHGPEKSEKQFRTFTIGKKEKSHS 125			
QY	135 YVITATSDTRGSLGSGVCLTRGMKYLAVGOSPRGAVPRKPVSEMERD----- 189			
DB	126 YVYISKPIYHQE-----TQCLKLVKTVNGKITSPHAYVQ-----EKQLQADDEPVQ 174			
QY	190 --RGAHSLP 198			
DB	175 VLSHIGSHAAP 165			
FT	RESULT 25			

EPH1\_MOUSE

ID	EPH1_MOUSE	STANDARD	PRT	205 AA
AC	P52793; P97331			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)			
DE	(LEKX-1) (Immediate early response protein B61)			
GN	EPHNA1 OR EPHL1 OR LERK1 OR EPL1			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=ICR;			
RX	MEDLINE=95405853; PubMed=7675446;			
RA	Takahashi H., Ikeda T.;			
RT	"Molecular cloning and expression of rat and mouse B61 gene;			
RT	implications on oncogenesis.";			
RL	Oncogene 11:879-883(1995).			
CC	-1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,			
CC	EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1.			
CC	(BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By			
CC	similarity).			
CC	-1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D38146; BAA07344.1; -			
DR	EMBL; U26188; AAB67563.1; -			
DR	EMBL; U90662; AAB50237.1; -			
DR	MED; MG1; 103236; Efnal.			
DR	InterPro: IPR001799; Ephrin.			
DR	Pfam: PF00812; Ephrin; 1.			
DR	PRINTS; PR01347; EPHRIN.			
DR	ProDom; PD002533; Ephrin; 1.			
DR	PROSITE; PS01299; EPHRIN; 1.			
KW	Glycoprotein; GPI-anchor; Signal.			
FT	SIGNAL 1 17			
FT	CHAIN 18 205			
FT	CARBOHYD 26 205			
FT	CONFLICT 74 74			
FT	CONFLICT 79 79			
FT	CONFLICT 81 81			
FT	CONFLICT 91 91			
FT	CONFLICT 94 94			
FT	CONFLICT 112 112			
FT	CONFLICT 115 115			
FT	CONFLICT 138 138			
FT	CONFLICT 154 154			
FT	CONFLICT 156 156			
FT	CONFLICT 159 159			
FT	CONFLICT 161 161			
CC	N-LINKED (GLCNAC... ) (POTENTIAL).			
CC	-----			
CC	Query Match 6.6%; Score 162.5; DB 1; Length 205;			
CC	Best Local Similarity 25.7%; Pred. No. 0.0003;			
CC	Matches 49; Conservative 33; Mismatches 66; Indels 23; Gaps 6;			
CC	-----			
QY	18 LIGLVGLVSGSLPEPYMNSAKRFOAEGGYLYFOIGRDLDPKRRAPGPHSSPYE 77			
DB	8 LIGLCCSLAIVRHRIVFMWNSNPKFRRE-DYTVHQLNDYDILICPHYEDDSV-ADAAME 65			
QY	78 FYKLYLVGAQGRCEAPAPNILLTCRPLD---DLFTIFQCEYSENLWGHEFRSHD 134			
DB	66 RYSLVYVHEQEVYTCBPQSKQVAKNCPKSAKHGPEKSEKQFRTFTIGKKEKSHS 125			
QY	135 YVITATSDTRGSLGSGVCLTRGMKYLAVGOSPRGAVPRKPVSEMERD----- 189			
DB	126 YVYISKPIYHQE-----TQCLKLVKTVNGKITSPHAYVQ-----EKQLQADDEPVQ 174			
QY	190 --RGAHSLP 198			
DB	175 VLSHIGSHAAP 165			
FT	RESULT 25			





RA Garofalo S., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse alpha 2(I) and alpha 1(X)  
 RT collagen mRNAs."  
 RL Biochim. Biophys. Acta 1130:78-80(1992).  
 CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC  
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- PTM: POLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isdb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X67348; CAA47763.1; -  
 DR EMBL; X65121; CAA46237.1; -  
 DR EMBL; X63013; CAA44741.1; -  
 DR EMBL; Z21610; CAA79736.1; -  
 DR PIR; S28807; S28807.  
 DR PIR; S31216; S31216.  
 DR PIR; S22215; S22215.  
 DR MGI; 88445; Col10a1.  
 DR InterPro; IPR001073; C1q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR PRODom; PD000007; Collagen; 2.  
 DR SMART; SM00110; C1Q; 1.  
 DR ProSITE; PS01113; C1Q; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cartilage; Collagen; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 1 18  
 FT DOMAIN 19 680  
 FT DOMAIN 57 56  
 FT DOMAIN 57 519  
 FT DOMAIN 520 680  
 FT DOMAIN 545 680  
 FT DOMAIN 248 248  
 FT CONFLICT 286 286  
 FT CONFLICT 306 306  
 FT CONFLICT 417 417  
 FT CONFLICT 451 451  
 FT CONFLICT 500 500  
 FT CONFLICT 567 567  
 FT CONFLICT 569 569  
 FT CONFLICT 571 571  
 FT CONFLICT 635 635  
 SQ SEQUENCE 680 AA; 66775 MW; FE984CA9AFA708E2 CRC64;  
 Query Match 6.5%; Score 159; DB 1; Length 680;  
 Best Local Similarity 25.5%; Pred. No. 0.0017;  
 Matches 97; Conservative 30; Mismatches 108; Indels 146; Gaps 24;  
 QY 1 MGPPHSGPGV-RVGAALLLGLVGLSLPVPVNSANKRQAGGYLVLPQIGDRID 59  
 DB 211 IGPP--GPSVGRRGEGFPGQPGI-----KGRGPPGEMG----- 244  
 QY 60 LDCPRARPPHSPNNEFKYLIVNGAGRCRCEAPRPLLTORPDLRFTIKQE 119  
 DB 245 ----PSGPPRPGQPP-----GKQGR--EGIGKPEALGSPQPGI----- 277  
 QY 120 YSPNLWGHERRSHDYIATSDGT---REGLEIQGVCLTRGKVLRLVQGSFRGA 175  
 DB 278 --PQEKHGPAGP-----IAGPQAPQFGKQGLPGLRG-----QRG-----PAG-- 314

QY 176 VPKRVSSEMERDGAHSLSEFKENLPDPTSNATSRGAEPLPPSPMPAVAGAAGL 235  
 DB 315 LPGAPG-----KSGRPAG--HPGEGPLPSF-----GNMGPGPKIPGNHGIPIGAKGEI 364  
 QY 236 ALLLLGVAGAGMCWRRAKP-----SESHRG-----PSGFRGGSLGLGGGG 282  
 DB 365 G-LVSGAPPGA--RGAPPGLDKGTYPGEPGLNPKNPLPQKAGDPVGATPG 419  
 QY 283 M-----GPREAPGELGIALRGGAADPPCFHYEKVSGDYGHVY 323  
 DB 420 LRGPVGVAKGVGNHGEAGPR-GERGIRGTR---GPTGPPGVDPFGSKGDPGNP-- 472  
 QY 324 IVDDGP-----PQSP 334  
 DB 473 -GAPPGAGIATKGLNGEPTGP 492  
 RESULT 28  
 ID CAL3\_BOVIN STANDARD, PRT, 1049 AA.  
 AC P04258;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain.  
 GN COL3A1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 CX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE OF 1-242.  
 RX MEDLINE=80026026; PubMed=488906;  
 RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wächter E.,  
 RA Kühn K.;  
 RT "The covalent structure of calf skin type III collagen. I. The amino  
 RT acid sequence of the amino terminal region of the alpha 1(III) chain  
 RT (positions 1-222).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).  
 RN [2]  
 RP SEQUENCE OF 243-422.  
 RX MEDLINE=80026027; PubMed=488907;  
 RA Dewes H., Fietzek P.P., Kühn K.;  
 RT "The covalent structure of calf skin type III collagen. II. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2  
 RT (positions 223-402).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).  
 RN [3]  
 RP SEQUENCE OF 423-571.  
 RX MEDLINE=80026028; PubMed=488908;  
 RA Bentz H., Fietzek P.P., Kühn K.;  
 RT "The covalent structure of calf skin type III collagen. III. The  
 RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4  
 RT (positions 403-551).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).  
 RN [4]  
 RP SEQUENCE OF 572-808.  
 RX MEDLINE=80026029; PubMed=488909;  
 RA Lang H., Glanville R.W., Fietzek P.P., Kühn K.;  
 RT "The covalent structure of calf skin type III collagen. IV. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5  
 RT (positions 552-788).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).  
 RN [5]  
 RP SEQUENCE OF 809-947.  
 RX MEDLINE=80026030; PubMed=488910;  
 RA Dewes H., Fietzek P.P., Kühn K.;  
 RT "The covalent structure of calf skin type III collagen. V. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A  
 RT (position 789-927).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).  
 RN [6]



RP SEQUENCE OF 948-1049.  
 RX MEDLINE=80026031; PubMed=488911;  
 RA Allman H., Fietzek P.P., Glanville R.W., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. VI. The amino  
 RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha  
 RT 1(III)CB98 (positions 928-1028).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).  
 CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC -1- ALONG WITH TYPE I COLLAGEN.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLISINES.  
 CC -1- PFM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 DR PIR; A02862; CGB075.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001007; VWF C.  
 DR Pfam; PF01391; Collagen;\_T7.  
 DR ProDom; PD000007; Collagen;\_1.  
 DR PROSITE; PS01208; VWF; EPRITL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Collagen.  
 FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1041 1049 NONHELICAL REGION (C-TERMINAL).  
 FT MOD\_RES 95 95 HYDROXYLATION.  
 FT MOD\_RES 107 107 HYDROXYLATION.  
 FT MOD\_RES 119 119 HYDROXYLATION.  
 FT MOD\_RES 938 938 HYDROXYLATION.  
 FT MOD\_RES 950 950 HYDROXYLATION.  
 FT CARBOHYD 107 107 O-LINKED (GAL. . .).  
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).  
 FT DISULFID 1040 1040 INTERCHAIN.  
 FT DISULFID 1041 1041 INTERCHAIN.  
 SQ SEQUENCE 1049 AA; 93651 MW; 8EEC33D1C66C9A3 CRC64;  
 Query Match 6.3%; Score 154.5; DB 1; Length 1049;  
 Best Local Similarity 26.5%; Pred. No. 0.0054;  
 Matches 60; Conservative 9; Mismatches 74; Indels 83; Gaps 8;  
 QY 171 PRGAVPRKPYSEMERDRGA-----ASHLEPKEN 202  
 DB 688 PAGSGPAGPFGQVKGERSPGPGAGPFGGPPGSGNGNGPSSGAPGKDG 747  
 QY 203 LRGPTNAT-----SRGAEPLPPSPMAVAGAGLALLLVAGA 245  
 DB 748 PRGPGSGAGSPGIGPKGDSGPPGERAPGPGPPGAPGLGAG-----LTARGL 802  
 QY 246 GGACWRRRRRAKP-----SSSRHPPG-----GSPFGSGSLGCGGCGPREAPGE 292  
 DB 803 AGPGWPGARSPGPGIKENGKPGSGNGERGPFGGLGAGTAGEPDRDGNPDS 862  
 QY 293 LGILRG-----GADPPPCPHYKVSQYGHVYIVQDGP 330  
 DB 863 DGLRGDAPGAPKADKRGNGSPGAP-----GAPGP-----GPP 896  
 RESULT 29  
 CC02 CAEEL STANDARD; PRT; 301 AA.  
 AC P17656;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cuticle collagen 2 precursor.  
 GN COL-2 OR M01B6.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pseudocercariae; Caenorhabditis.  
 OC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;

RX MEDLINE=83050944; PubMed=7139711;  
 RA Kramer J.M., Cox G.N., Hirsch D.;  
 RT "Comparisons of the complete sequences of two collagen genes from  
 RT Caenorhabditis elegans.";  
 RL Cell 30:559-606(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=85105075; PubMed=2578467;  
 RA Kramer J.M., Cox G.N., Hirsch D.;  
 RT "Expression of the Caenorhabditis elegans collagen genes col-1 and  
 RT col-2 is developmentally regulated.";  
 RL J. Biol. Chem. 260:1945-1951(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Coles L.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; J01048; AAA27990.1; -;  
 DR EMBL; V00148; CAA23464.1; -;  
 DR EMBL; Z68301; CAA92620.1; -;  
 DR PIR; B31219; B31219.  
 DR WormPep; W0186.7; CE03759.  
 DR InterPro; IPR002486; COL\_cuticle\_N.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen;\_2.  
 DR Pfam; PF01484; COL\_cuticle\_N;\_1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; collagen;  
 KM Signal.  
 FT SIGNAL 1 37 POTENTIAL.  
 FT CHAIN 38 301 CUTICLE COLLAGEN 2.  
 FT DOMAIN 85 97 GLY-RICH.  
 FT DOMAIN 105 134 TRIPLE-HELICAL REGION.  
 FT DOMAIN 153 176 TRIPLE-HELICAL REGION.  
 FT DOMAIN 183 212 TRIPLE-HELICAL REGION.  
 FT DOMAIN 215 282 TRIPLE-HELICAL REGION.  
 SQ SEQUENCE 301 AA; 28025 MW; 33317E3BDAC0229 CRC64;  
 Query Match 6.2%; Score 151.5; DB 1; Length 301;  
 Best Local Similarity 28.7%; Pred. No. 0.0024;  
 Matches 50; Conservative 15; Mismatches 72; Indels 37; Gaps 7;  
 QY 168 GQPRGAVPRKPYSEMERDRG-----ASHLEPKENPGDPT--SNATSRAG 217  
 DB 129 GNPFGAGAPCEPVTPQCPGPPGAPGAPGPPGPGSGAGSGGAPGPP 188  
 QY 218 GELPPSPMAVAGAGLALLLVAGA---GGACWRRRRRAKPSRRHPPG--GSPFGSG 273  
 DB 189 GPAGPAGNDAGAPGAPG-----PGSPGSEGGP-----GEPGAPAPGPPAGAGNDG 236  
 QY 274 SLGLGCGGGMGR-----EAPRGELIALRGGAADPPPCPHYKVS 316  
 DB 237 AGTGGPAPGAPGKPPGAPGAGADNGNPGGPGTAKPGGEGEKICPYCAIDG 290  
 RESULT 30  
 CA34\_HUMAN

ID CA34\_HUMAN STANDARD; PRT; 1670 AA.  
 AC Q01955; OSB012;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
 GN COL4A3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94364994; PubMed=8083201;  
 RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reekers S.T.;  
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
 RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
 RT human tissues.";  
 RL J. Biol. Chem. 269:23013-23017(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Leinonen A.;  
 RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;  
 RP Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;  
 RP E-163; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.  
 RX MEDLINE=21064696; PubMed=11334255;  
 RA Heidt L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
 RA Gutierrez B., Stavarou C., Gubler M.C., Antignac C.;  
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
 RT autosomal Alport syndrome.";  
 RL J. Am. Soc. Nephrol. 12:97-106(2001).  
 RN [4]  
 RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=93015826; PubMed=1400291;  
 RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Sans J.;  
 RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
 RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
 RT antigenic region at the triple helix/NC1 domain junction.";  
 RL J. Biol. Chem. 267:19780-19784(1992).  
 RN [5]  
 RP SEQUENCE OF 1453-1670 FROM N.A.  
 RX MEDLINE=91353570; PubMed=1882840;  
 RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reekers S.T.;  
 RT "Sequence and localization of a partial cDNA encoding the human alpha  
 RT 3 chain of type IV collagen.";  
 RL Am. J. Hum. Genet. 49:545-554(1991).  
 RN [6]  
 RP SEQUENCE OF 1331-1670 FROM N.A.  
 RX TISSUE=Kidney;  
 RX MEDLINE=92147878; PubMed=1737849;  
 RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
 RA Pusey C.D.;  
 RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
 RT to be the alpha 3 chain of type IV collagen.";  
 RL J. Clin. Invest. 89:592-601(1992).  
 RN [7]  
 RP SEQUENCE OF 1644-1670 FROM N.A.  
 RC TISSUE=Kidney;  
 RN Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94124597; PubMed=8294492;  
 RA Feng L., Xia Y., Wilson C.B.;  
 RT "Alternative splicing of the NCI domain of the human alpha 3(IV)  
 RT collagen gene. Differential expression of mRNA transcripts that  
 RT predict three protein variants with distinct carboxyl regions.";  
 RL J. Biol. Chem. 269:2342-2348(1994).  
 RN [9]

RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Momota R., Sugimoto M., Kohashi T., Kigasawa K., Yoshioke H.,  
 RA Nishimura Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
 RT 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [10]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=93280184; PubMed=8505332;  
 RA Bernal D., Quinones S., Sans J.;  
 RT "The human mRNA encoding the Goodpasture antigen is alternatively  
 RT spliced.";  
 RL J. Biol. Chem. 268:12090-12094(1993).  
 RN [11]  
 RP VARIANT PRO-1474.  
 RX MEDLINE=95078827; PubMed=7987301;  
 RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
 RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,  
 RA Reekers S.T., Smeets H.U.M.;  
 RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
 RT recessive Alport syndrome.";  
 RL Hum. Mol. Genet. 3:1269-1273(1994).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/  
 CC NIDDOEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: 1 (SHOWN HERE), 2/V AND  
 CC 3/L5 ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR  
 CC C-TERMINAL NCI DOMAINS.  
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL  
 CC N-LINKED GLYCOSYLATION SITE.  
 CC -1- PTM: TYPES IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
 CC -1- DISEASE: ANTIBODIES AGAINST THE NCI DOMAIN OF ALPHA3(IV) MEDIATE  
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS  
 CC CHARACTERIZED BY HEMATURIA AND PUTOKARY HEMORRAGE.  
 CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I  
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY  
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
 CC MALES AND FEMALES.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X80031; CA56335.1; -;  
 CC EMBL; AJ288487; CAC36101.1; -;  
 CC EMBL; AJ288488; CAC36101.1; JOINED.  
 CC EMBL; AJ288489; CAC36101.1; JOINED.

DR EMBL/ AJ288490; CAC36101.1; JOINED.  
 DR EMBL/ AJ288491; CAC36101.1; JOINED.  
 DR EMBL/ AJ288492; CAC36101.1; JOINED.  
 DR EMBL/ AJ288493; CAC36101.1; JOINED.  
 DR EMBL/ AJ288494; CAC36101.1; JOINED.  
 DR EMBL/ AJ288495; CAC36101.1; JOINED.  
 DR EMBL/ AJ288496; CAC36101.1; JOINED.  
 DR EMBL/ AJ288497; CAC36101.1; JOINED.  
 DR EMBL/ AJ288498; CAC36101.1; JOINED.  
 DR EMBL/ AJ288499; CAC36101.1; JOINED.  
 DR EMBL/ AJ288500; CAC36101.1; JOINED.  
 DR EMBL/ AJ288501; CAC36101.1; JOINED.  
 DR EMBL/ AJ288502; CAC36101.1; JOINED.  
 DR EMBL/ AJ288503; CAC36101.1; JOINED.  
 DR EMBL/ AJ288504; CAC36101.1; JOINED.  
 DR EMBL/ AJ288505; CAC36101.1; JOINED.  
 DR EMBL/ AJ288506; CAC36101.1; JOINED.  
 DR EMBL/ AJ288507; CAC36101.1; JOINED.  
 DR EMBL/ AJ288508; CAC36101.1; JOINED.  
 DR EMBL/ AJ288509; CAC36101.1; JOINED.  
 DR EMBL/ AJ288510; CAC36101.1; JOINED.  
 DR EMBL/ AJ288511; CAC36101.1; JOINED.  
 DR EMBL/ AJ288512; CAC36101.1; JOINED.  
 DR EMBL/ AJ288513; CAC36101.1; JOINED.  
 DR EMBL/ AJ288514; CAC36101.1; JOINED.  
 DR EMBL/ AJ288515; CAC36101.1; JOINED.  
 DR EMBL/ AJ288516; CAC36101.1; JOINED.  
 DR EMBL/ AJ288517; CAC36101.1; JOINED.  
 DR EMBL/ AJ288518; CAC36101.1; JOINED.  
 DR EMBL/ AJ288519; CAC36101.1; JOINED.  
 DR EMBL/ AJ288520; CAC36101.1; JOINED.  
 DR EMBL/ AJ288521; CAC36101.1; JOINED.  
 DR EMBL/ AJ288522; CAC36101.1; JOINED.  
 DR EMBL/ AJ288523; CAC36101.1; JOINED.  
 DR EMBL/ AJ288524; CAC36101.1; JOINED.  
 DR EMBL/ AJ288525; CAC36101.1; JOINED.  
 DR EMBL/ AJ288526; CAC36101.1; JOINED.  
 DR EMBL/ AJ288527; CAC36101.1; JOINED.  
 DR EMBL/ AJ288528; CAC36101.1; JOINED.  
 DR EMBL/ AJ288529; CAC36101.1; JOINED.  
 DR EMBL/ AJ288530; CAC36101.1; JOINED.  
 DR EMBL/ AJ288531; CAC36101.1; JOINED.  
 DR EMBL/ AJ288532; CAC36101.1; JOINED.  
 DR EMBL/ AJ288533; CAC36101.1; JOINED.  
 DR EMBL/ AJ288534; CAC36101.1; JOINED.  
 DR EMBL/ AJ288535; CAC36101.1; JOINED.  
 DR EMBL/ AJ288536; CAC36101.1; JOINED.  
 DR EMBL/ AJ288537; CAC36101.1; JOINED.  
 DR EMBL/ AJ288538; CAC36101.1; JOINED.  
 DR EMBL/ M92993; AAA21610.1; -.  
 DR EMBL/ S55790; AAB19637.1; -.  
 DR EMBL/ M81379; AAA51556.1; -.  
 DR EMBL/ L08650; AAA52044.1; -.  
 DR EMBL/ U02519; AAA18942.1; -.

Query Match 6.1%; Score 148.5; DB 1; Length 1670;  
 Best Local Similarity 27.1%; Pred. No. 0.022;  
 Matches 105; Conservative 22; Mismatches 146; Indels 115; Gaps 23;

QY 1 MGPHS-GPGGV-----RVGALLILGLVLSLEPYVYVNSANKRFOAGGYLYPQI 54  
 DB 1020 MWMGSKKKGKGTGLGFGRAARPGIPIHGL-QDDKGPPI-----SEG----- 1061  
 QY 55 GDRIDLICPARPPGPHSSPNVEPYKYLIVGAGRRCC-----APPA 97  
 DB 1062 -----TPPGPGPTGDDG-----LPDMGKKGEWGQPGPGHLPAGPEGAPGS 1105  
 QY 98 PNL--ILTCORPDLRLFT--IKPQEVSPNLWHE---FRSHDYIITASDSTRGLS 150  
 DB 1106 PGRGLGKGFHDDLGFKIKGLGPRPGIRPGPLGFGSGPGKIRKQDQ-RGIRG 1164  
 QY 151 LGGGVLITRGMKVLILRVGQSPRG--GAVPRKPVSEMPMERDRGAASLEBGENLPGDPT 208

DB 1165 PAG---EKETGLIRAPPGRGNPDGAQAK-----GRCGA-----PGFPLPG--- 1204  
 QY 209 SNATSRGAEGPLPEPSPMBAVAGAGLALLILGVAGAGAMCMRRARRKAPSESRHPG-- 266  
 DB 1205 -RKAGMAGAPRGPTGIGFPGPPLPGAILPGQGNRGP---PGRSGSPGAPPGPPG 1260  
 QY 267 -----GTFGRGSLG-LGGGGGKGP--REAPPELIALRGGAAPPFCPIHEK 313  
 DB 1261 SHVIGIKDDKSGSMGHPGKPPGTAGDMGPPGRLGAPGTPLP---GPRGDPGF-QGFP 1316  
 QY 314 VSGDYGHPVYIVODGP-----POSPPNI 336  
 DB 1317 VKGEKGNPFLGSLGPPPIPGKPPGV 1344

RESULT 31  
 ID CAH MOUSE STANDARD; PRT; 1527 AA.  
 AC P39061; 062002; 061437;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].  
 GN COL18A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=94245707; PubMed=8186673;  
 RA Rehn M.V., Hirtikka E., Pihlajantani T.;  
 RT "Primary structure of the alpha 1 chain of mouse type XVII collagen,  
 RT partial structure of the corresponding gene, and comparison of the  
 RT alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen  
 RT chain.";  
 RL J. Biol. Chem. 269:13929-13935(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).  
 RA Rehn M., Hirtikka E., Pihlajantani T.;  
 RT "Characterization of the mouse gene for the alpha-1 chain of type  
 RT XVII collagen (COL18A1) reveals that the three variant N-terminal  
 RT polypeptide forms are transcribed from two widely separated  
 RT promoters.";  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).  
 RX MEDLINE=94240112; PubMed=8183894;  
 RA Rehn M.V., Pihlajantani T.;  
 RT "Alpha 1(XVII), a collagen chain with frequent interruptions in the  
 RT collagenous sequence, a distinct tissue distribution, and homology  
 RT with type XV collagen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4236(1994).  
 RN [4]  
 RP SEQUENCE OF 240-1527 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94240111; PubMed=8183893;  
 RA Oh S.P., Kamagata Y., Murgazaki Y., Timmons S., Ooshima A., Olsen B.R.,  
 RT "Isolation and sequencing of cDNAs for proteins with multiple domains  
 RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous  
 RT proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).  
 RN [5]  
 RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.  
 RX MEDLINE=97160848; PubMed=9008166;  
 RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,  
 RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;  
 RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor  
 RT growth.";  
 RL Cell 88:277-285(1997).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.

RX MEDLINE=98169382; PubMed=9501087;  
 RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;  
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A  
 resolution."  
 RL EMBL J. 17.1656-1664(1998).  
 CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
 CC FACTOR SIGNALLING.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM  
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLETT REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L16898; AAA37434.1; -;  
 DR EMBL; U03715; AAC22901.1; -;  
 DR EMBL; U03715; AAC22901.1; JOINED.  
 DR EMBL; U34606; AAC22901.1; JOINED.  
 DR EMBL; U34608; AAC22901.1; JOINED.  
 DR EMBL; U34609; AAC22901.1; JOINED.  
 DR EMBL; U34610; AAC22901.1; JOINED.  
 DR EMBL; U34611; AAC22901.1; JOINED.  
 DR EMBL; U34612; AAC22901.1; JOINED.  
 DR EMBL; U34613; AAC22901.1; JOINED.  
 DR EMBL; U03716; AAC22901.1; JOINED.  
 DR EMBL; U03718; AAC22901.1; JOINED.  
 DR EMBL; U03715; AAC22901.1; -;  
 DR EMBL; U34607; AAC22902.1; JOINED.  
 DR EMBL; U34608; AAC22902.1; JOINED.  
 DR EMBL; U34609; AAC22902.1; JOINED.  
 DR EMBL; U34610; AAC22902.1; JOINED.  
 DR EMBL; U34611; AAC22902.1; JOINED.  
 DR EMBL; U34612; AAC22902.1; JOINED.  
 DR EMBL; U34613; AAC22902.1; JOINED.  
 DR EMBL; U03716; AAC22902.1; JOINED.  
 DR EMBL; U03718; AAC22902.1; JOINED.  
 DR EMBL; U11636; AAC22178.1; -;  
 DR EMBL; L22545; AAA19767.1; -;  
 DR PDB; 1KOE; 16-FEB-99.  
 DR MGD; MGI:88451; Coll18a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001791; Lamlnin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 8.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
 KW 3d-structure.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1527  
 FT CHAIN 1344 1527  
 FT DOMAIN 27 538  
 FT DOMAIN 539 555  
 FT DOMAIN 556 575  
 FT DOMAIN 576 649  
 FT DOMAIN 650 673  
 FT DOMAIN 674 795  
 FT DOMAIN 796 818  
 FT DOMAIN 819 901  
 FT DOMAIN 902 915  
 FT DOMAIN 916 957  
 POTENTIAL.  
 COLLAGEN ALPHA 1 (XVIII) CHAIN.  
 ENDOSTATIN.  
 NONHELICAL REGION 1 (NC1).  
 TRIPLE-HELICAL REGION 1 (COL1).  
 NONHELICAL REGION 2 (NC2).  
 TRIPLE-HELICAL REGION 2 (COL2).  
 NONHELICAL REGION 3 (NC3).  
 TRIPLE-HELICAL REGION 3 (COL3).  
 NONHELICAL REGION 4 (NC4).  
 TRIPLE-HELICAL REGION 4 (COL4).  
 NONHELICAL REGION 5 (NC5).  
 TRIPLE-HELICAL REGION 5 (COL5).

FT DOMAIN 958 970  
 FT DOMAIN 971 1043  
 FT DOMAIN 1044 1053  
 FT DOMAIN 1054 1086  
 FT DOMAIN 1087 1098  
 FT DOMAIN 1099 1122  
 FT DOMAIN 1123 1129  
 FT DOMAIN 1130 1181  
 FT DOMAIN 1182 1194  
 FT DOMAIN 1195 1212  
 FT DOMAIN 1213 1527  
 FT CARBOHYD 338 338  
 FT CARBOHYD 700 700  
 FT DISULFID 1376 1516  
 FT DISULFID 1478 1508  
 FT SITE 1104 1106  
 FT SITE 1 212  
 FT VARSPLIC 213 238  
 FT FT  
 FT CONFLICT 900 900  
 FT CONFLICT 947 947  
 FT CONFLICT 964 964  
 FT CONFLICT 1157 1157  
 FT CONFLICT 1266 1266  
 FT CONFLICT 1276 1276  
 FT CONFLICT 1437 1437  
 SQ SEQUENCE 1527 AA; 156008 MW; 9645045AF140B513 CRC64;  
  
 Query Match 6.0%; Score 146.5; DB 1; Length 1527;  
 Best Local Similarity 23.7%; Pred. No. 0.027; Indels 167; Gaps 21;  
 Matches 92; Conservative 22; Mismatches 107;  
 QY 2 GPHSGPGVGVGALLLVGLVGLVSGLSLEPYVWNSANKRFQAEQGVLYLPQIGRLDL 60  
 DB 551 GPP--GPAGPQ-----GPAGPVVQSPNSQPV----- 574  
 QY 61 LCPRAR-PPGHSSPNVEFYKLYLVGAQGRCE----- 93  
 DB 575 --PQAGPFPGPQPPGPGD-----GTFGRDDEPDGPDGPDGTGFGPGRPDGVD 624  
 QY 94 -----APPANLLTCDRBDLARTIRFGQSVSPNLMGHRSHHDYLIAT 140  
 DB 625 PKRKGDPGIGPPGPPGP-----PPPPSPSRDKLTITDME 661  
 QY 141 SDGTREGESLQGVCLTRGMKVLIVGQSPRGCAVPRRP--VSENMKMRDR-GAAHSLE 197  
 DB 662 GSGFSGDIESLRG-----PRGFPDPGPPGVGGLPGEPRGINGSYA 704  
 QY 198 PKENLPDPTSNATSRGARGP--LP-PPSMRAVAGAAGLALLLGVAGAGAMCMRRR 254  
 DB 705 PGPAGLPGPV-----GKRGPPGFPGPPGPPPKRGP-----DGVAGQKQSV----- 747  
 QY 255 RAKPSRSRHPGP-GSPFGGSLGLGGGKGWGPRAEPGELIALRGG--GAADPFCPHY 311  
 DB 748 ----GDVGIRPGKSGKCDLGPIMPGKSGLA---GSPGPVGPFGPPGPPGPPGPAAGF 800  
 QY 312 EKVSQDYGHFVTVQ-----DPPQSP 333  
 DB 801 DMEGSG-GIPLWTTARSSDGLQGPFPSP 827  
  
 RESULT 32  
 CAFI\_RIFPA  
 ID CAFI\_RIFPA STANDARD; PRT; 1027 AA.  
 AC P30754;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE fibril-forming collagen alpha chain.  
 OS Riffia pachyptila (Tube worm).  
 OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Acanthobranchia;  
 OC Riffiada; Riffiidae; Riffia.  
 OX NCBI\_TaxID=6426;

RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=93130909; Pubmed=1483468;  
 RA Mann K., Gallil F., Timpl R.;  
 RT "Amino-acid sequence and cell-adhesion activity of a fibril-forming  
 RT collagen from the tube worm *Riftia pachyptila* living at deep sea  
 RT hydrothermal vents.";  
 RL Eur. J. Biochem. 210:839-847(1992).  
 RN [2]  
 RP SEQUENCE OF 8-45; 525-618 AND 810-882.  
 RC TISSUE=Cuticle;  
 RX MEDLINE=92015209; Pubmed=1920405;  
 RA Gallil F., Wiedemann H., Mann K., Kuhn K., Timpl R., Engel J.;  
 RT "Molecular characterization of cuticle and interstitial collagens  
 RT from worms collected at deep sea hydrothermal vents.";  
 RL J. Mol. Biol. 221:209-223(1991).  
 CC -1- FUNCTION: FIBRIL-FORMING COLLAGEN.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- PTM: ALL THE "X" ARE PROBABLY GLYCOSYLATED HYDROXYLISINES.  
 DR PIR; S22915; S22915.  
 DR Interpro; IPR00087; Collagen.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Collagen.  
 FT DOMAIN 1 12 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 13 1023 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1024 1027 NONHELICAL REGION (C-TERMINAL).  
 FT MOD\_RES 21 21 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 24 24 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 27 27 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 39 39 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 53 53 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 54 54 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 72 72 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 90 90 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 93 93 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 123 123 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 128 128 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 150 150 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 161 161 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 162 162 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 164 164 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 165 165 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 174 174 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 177 177 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 180 180 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 183 183 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 207 207 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 216 216 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 219 219 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 228 228 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 237 237 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 243 243 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 249 249 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 255 255 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 273 273 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 276 276 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 285 285 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 291 291 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 303 303 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 306 306 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 312 312 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 321 321 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 327 327 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 339 339 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 342 342 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 348 348 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 351 351 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 366 366 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 372 372 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 375 375 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 381 381 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 387 387 HYDROXYLATION (PARTIAL).

FT MOD\_RES 416 416 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 417 417 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 423 423 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 429 429 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 432 432 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 433 433 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 453 453 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 465 465 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 483 483 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 500 500 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 503 503 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 506 506 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 513 513 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 525 525 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 533 533 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 536 536 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 540 540 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 546 546 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 551 551 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 552 552 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 561 561 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 603 603 IMPERFECTION IN THE GAA REPEAT.  
 FT SITE 610 610 IMPERFECTION IN THE GAA REPEAT.  
 FT MOD\_RES 621 621 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 627 627 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 645 645 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 647 647 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 648 648 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 663 663 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 708 708 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 711 711 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 714 714 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 717 717 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 723 723 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 744 744 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 759 759 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 773 773 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 774 774 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 783 783 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 792 792 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 815 815 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 816 816 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 843 843 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 849 849 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 855 855 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 861 861 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 867 867 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 888 888 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 894 894 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 903 903 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 915 915 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 933 933 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 939 939 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 945 945 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 954 954 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 963 963 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 966 966 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 984 984 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 990 990 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1010 1010 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1011 1011 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1013 1013 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1014 1014 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1016 1016 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1017 1017 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1019 1019 HYDROXYLATION (PARTIAL).  
 FT MOD\_RES 1020 1020 HYDROXYLATION (PARTIAL).  
 FT VARIANT 903 903 P -> A.  
 SQ SEQUENCE 1027 AA; 94353 MM; 9B043FA7788386C1 CRC64;

Query Match 6.0%; Score 146; DB 1; Length 1027;  
 Best Local Similarity 22.2%; Pred. No. 0.019;  
 Matches 82; Conservative 24; Mismatches 113; Indels 150; Gaps 17;

```

QY 46 GGVLYPQIGDRDLDCPPRAPPSPHSPNVEFYKLYLVGAQGRCE----- 93
DB 4 GPRVYIAQVGP-----IGPRG-PPGPRGSPGQ-----GYGLRGERGDSGMPKPIGK 50
QY 94 -APPANLLTCDRDLRFLTKQVYSNLMGHEFRSHHYYIANSDBRGLBSLQ 152
DB 51 RGPGR-----AGIAG-----KSGDDRDDEPGR 75
QY 153 GGVCLT-----RGMKVL-----LRVQSPRG--AVPRKPVSEM 184
DB 76 GGIQVPGPRGAGGMPGMPGXGHRGPRGSGSGXGEGKSGNGGPGGPPAPSPGPIGR 135
QY 185 PHERDGAHSLERPEKENTPGDPTSNATSRGAEGPLPPSPMAVAGAGCLAL----- 238
DB 136 GGTGERG-----RDGKSGLPF-----LRVDGLAGPPGPGPIGSTSPGPRPGR 183
QY 239 -----LLGVAGAGGAGMCMRRBRRAKPSRSPG-PGSFGRGSLGLGGGGM----- 283
DB 184 GDRGSGTGAQGLQGPVLSGQPGVAGENGHRGMPGMGANGEPGASGSLPPSGPR 243
QY 284 GPR-----EAPGELGIALRG-----GAADPPCPHYEKVSGDYGHPLYT 324
DB 244 GPRGMPGTAGSPGQAGXGCGPTGEGRPGAQVYAGSSGPP-----GDVAGFCHA 294
QY 325 VQDPPQSP 333
DB 295 GEAGKRGSP 303

RESULT 33
CA1H_BOVIN STANDARD; PRT; 674 AA.
ID CA1H_BOVIN STANDARD; PRT; 674 AA.
AC P23206;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=9111311; PubMed=1703407;
RA Thomas J.T., Kwan A.P.L., Grant M.B., Scott-Handford R.P.;
RT "Isolation of cDNAs encoding the complete sequence of bovine type X
RT collagen. Evidence for the condensed nature of mammalian type X
RT collagen genes.";
RT Biochem J 273:141-148 (1991).
CC - FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC - SUBUNIT: HOMOTRIMER.
CC - PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC - SIMILARITY: STRONG TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC - SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53556; CAA37624.1; -.
CC F013301; S13301.
CC InterPro; IPR001073; C1Q.
CC InterPro; IPR000087; Collagen.

```

```

DR Pfam; PF00386; C1Q; 1.
DR Pfam; PF01391; Collagen; 9.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR Prodom; PD000007; Collagen; 1.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cartilage; Collagen; Signal; Glycoprotein.
FT SIGNAL 1
FT CHAIN 19 674 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 56 NONHELICAL REGION (NC2).
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
FT DOMAIN 520 674 NONHELICAL REGION (NC1).
FT DOMAIN 539 674 C1Q.
FT DISULFID 194 197 BY SIMILARITY.
FT MOD RES 460 460 HYDROXYLATION (BY SIMILARITY).
FT MOD RES 463 463 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 674 AA; 65546 MW; CD4CA73A03E004CA CRC64;

Query Match 5.9%; Score 145; DB 1; Length 674;
Best Local Similarity 24.2%; Pred. No. 0.014;
Matches 100; Conservative 30; Mismatches 125; Indels 158; Gaps 24;

QY 15 ALLILGVTLGVSLERPEYVNSANKRFQEGGVLYPQIGDRDLDCPPA----- 65
DB 6 ALLILMSLNLVHG-----VRY-----TERVQPTG-IGPSPNTQTFPIPAIKGVSLR 56
QY 66 -----RPPGR-----HSSPNYEFYKLYLVG-----GAQR-RCEAPPANLLT 103
DB 57 GEGGIPGPPGAPGRCHGSP-----GPRKPTGSPGQGPGLPGRPPS--A 104
QY 104 CDRPDLRFTIKQVSPNLMGHEFRSHHYYIATSDGTRGLBSLQ----- 153
DB 105 TKGKGL-----PGLPGKQ-----GER-GLNGPKKIDIPAGLPGR 137
QY 154 -----GVCLTRGMKYLVRVG--QSPRGAVPR-----KEVSEMP-----MERDGAHSL 196
DB 138 RGPGRPGIPGPAGISVPGKRGQPGFGEPRGFPEKEKTSVPGINGQKGEHGCTPC 197
QY 197 EPGENIPGDPSTNATSRGAEGPLPP-----SNPANAAGAGLALLLVAGAGCA 248
DB 198 RFGGRGLPFG-----DQGTGPPGPPGVAKRGENGILPGPGKXGD-----QGVPERGA 245
QY 249 MCMRRRAKPSRSP-----GPSFGRGSLGLGGGGMGRAPAPDELIA----- 296
DB 246 -----AGSPGPGPRGEGPBGIGKPAIGQPGIGMKGQPGAPATAGAPAGF 298
QY 297 -----LRG-GGAADPPCPHYEKVSGDYGHPLYVQDPP-----PQSPNTI 336
DB 299 GKPGLPGLKGRGQHVGLPGSPGAKGEGGPAHGPEAGLPQSGNMGPQGGKI 351

RESULT 34
CA1H_HUMAN STANDARD; PRT; 1516 AA.
ID CA1H_HUMAN STANDARD; PRT; 1516 AA.
AC P39060; Q9Y608; Q9Y607; Q9UK38;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9616406; PubMed=9503365;
RA Sareela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT "Complete primary structure of two variant forms of human type XVIII
RT collagen and tissue-specific differences in the expression of the
RT corresponding transcripts.";

```



